

Imprecise Probability Models for Logistic Regression

A Thesis

Submitted to the College of Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

in the

Department of Mathematics and Statistics

University of Saskatchewan

Saskatoon

By

Osama Mohammad Bataineh

©Osama Mohammad Bataineh, September/2012. All rights reserved.

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Mathematics and Statistics
University of Saskatchewan,
Saskatoon, Saskatchewan, Canada
S7N 5E6

ABSTRACT

Imprecise probability models are applied to logistic regression to produce interval estimates of regression parameters. The lengths of interval estimates are of main interest. Shorter interval estimates correspond to less imprecision in regression parameters estimates.

This thesis applies imprecise probabilistic methods to the logit model. Imprecise logistic regression, briefly called ImpLogit model, is presented and established for the first time. ImpLogit model is applied based on an inferential paradigm that applies Bayes theorem to a family of prior distributions, yielding interval posterior probabilities. The so-called interval estimates of regression parameters are computed using Metropolis-Hastings algorithm.

Two imprecise prior probability models are applied to 2-parameter ImpLogit model : the imprecise Dirichlet model (IDM) and the imprecise logit-normal model (ILnM). The 2-parameter ImpLogit model is fitted using real life dose-response data. This takes into account the cases of increasing, decreasing and mixed-belief ImpLogit models.

The relation between the lengths of interval estimates of regression parameters and both of covariate values and imprecise prior hyperparameters, in 2-parameter ImpLogit model, is studied by simulation. Different designs are applied in order to investigate a way to shorten the lengths of interval estimates of regression parameters. Having covariate fixed values to surround the prior believed median value of the logistic distribution results in reducing the imprecision in interval estimates. Fixing covariate values around the prior believed median value in a short range increases the lengths of interval estimates.

The number of fixed covariate values (say number of distinct dose levels in a dose-response experiment) affects the produced imprecision. A larger number of fixed covariate values increases the lengths of interval estimates. Therefore, a good design has a small number of fixed covariate values, located and spread out not in a short range.

ImpLogit model designs that are recommended by the simulation study, are compared to optimal designs in the frequentist approach using Fisher information matrix (FIM). Designs in FIM agree with designs that reduce imprecision in 2-parameter ImpLogit model, in the necessity of having covariate values to be fixed around the prior believed

median value of the logistic distribution, not in a short range.

Keywords : Imprecise probability model; imprecise Dirichlet model; imprecise logit-normal model; aggregation property; ImpLogit model; interval estimate.

ACKNOWLEDGEMENTS

First, I would like to thank my supervisor, Dr. Mikelis Bickis for his encouragement, advice, enduring patience throughout the years of my study. His guidance has been vital and he has never been ceasing in his support in every possible way towards making me able, to accomplish my thesis. He has always provided clear explanations when I was stuck, constantly driving the research with new and innovative research methodologies.

I want to sincerely acknowledge the financial support provided by the College of Graduate Studies, Department of Mathematics and Statistics and Dr. Mikelis Bickis.

It is my great pleasure to thank all my friends and well-wishers who encouraged and helped me in the PhD program.

Finally, I am very much grateful to my parents, my sister and my younger brothers, for their encouragement during my stay far away from them.

To my parents whom I love the most ...

CONTENTS

Permission to Use	i
Abstract	ii
Acknowledgements	iii
Contents	vi
1 Introduction	1
1.1 Motivation	1
1.2 Logit Model	5
1.3 Bayesian Approach for the Logit Model	8
1.4 Bayesian Models Robustness	9
1.5 Historical Development of Bayesian Methods	11
1.5.1 Principle of Inverse Probability	12
1.5.2 Appearance of the Frequentist Approach	13
1.5.3 Establishment of the Subjective Bayesian Approach	13
1.6 Imprecise Probabilities	15
1.7 Imprecise Dirichlet Model	19
1.8 Applications of Imprecise Probability Models	25
1.9 Organization of the Thesis	27
2 Bayesian 2-Parameter Logit Model	28
2.1 Increasing 2-Parameter Logit Model	28
2.2 Decreasing 2-Parameters Logit Model	38
2.3 Logit–Normal Model	43
2.4 Multivariate Logit-Normal Model	45
3 ImpLogit Model	50
3.1 Imprecision in 2-Parameter Logit Model	51
3.2 Transmission from θ to β in 2-Parameter ImpLogit Model	61
3.2.1 Positive x_j and x_k	65
3.2.2 Negative x_j and x_k	67
3.2.3 Negative x_j and Positive x_k	68
3.3 2-Parameter Increasing ImpLogit Model	73
3.4 2-Parameter Decreasing ImpLogit Model	78
3.5 Mixture of Beliefs in 2-Parameter ImpLogit Model	82
3.6 ImpLogit with Imprecise Bivariate Logit-Normal Model	84
3.7 ImpLogit with Smaller Sample	87

4	Simulation Study	89
4.1	Motivation for the Simulation Study	89
4.2	Symmetric Two Values	91
4.3	Right and Left-Shifted Two Values	94
4.4	Extremely Shifted Two Values	95
4.5	Symmetric Covariate Multiple Values	97
4.6	Right and Left-Shifted Multiple Values	97
4.7	Extremely Shifted Multiple Values	98
4.8	Closer θ_j and θ_k	98
4.9	Optimal Frequentist Design	99
4.10	Multiple Parameter ImpLogit Model	102
5	Conclusion and Future Plan	130
5.1	Conclusion	130
5.2	Future Plan	133
	References	134
A	R Code	143
B	Posterior logit-normal density	150
C	Posterior multivariate logit-normal density	155
D	Calculations and graphs from Chapter 3	157
E	Graphs from Chapter 4	163
F	Fisher Information Matrix for Logistic Regression	171

CHAPTER 1

INTRODUCTION

1.1 Motivation

Logistic regression is a well known statistical model that has been applied widely in statistical analysis. It is used in applied statistics to analyze binary and multinomial data. In the logistic regression model (also called logit model), observed binary data are assumed to come from a Bernoulli distribution where the sum of observed binary data follows the binomial distribution. The logit model makes predictions about the probability of occurrence of an event. Finney (1978) used the logit model for applications in statistical bioassay in order to build dose-response relationships. Collet (1991) gives a detailed study of the logit model with applications to real life binary data. Fundamentals of analysis of binary data are given by Cox (1970) and Cox and Snell (1989). Applying the logistic regression model is discussed in Agresti (2007). A recent study for logistic regression model is given in Hilbe (2009).

The logit model assumes a random variable, say Y_i , $i = 1, \dots, m$, to follow the binomial distribution, $\text{bin}(n_i, \theta_i)$, where θ_i denotes the i th probability of success, and n_i is the number of trials in the i th observation. The main parameters of interest, θ_i for all i , are modelled as

$$\theta_i = \theta(\mathbf{x}_i) = \frac{e^{\mathbf{x}_i' \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i' \boldsymbol{\beta}}} = \frac{e^{\beta_0 + x_{i1}\beta_1 + \dots + x_{ip}\beta_p}}{1 + e^{\beta_0 + x_{i1}\beta_1 + \dots + x_{ip}\beta_p}},$$

$$-\infty < \beta_j < \infty, \quad j = 0, \dots, p,$$

$$-\infty < x_{ij} < \infty, \quad i = 1, \dots, m, \quad j = 1, \dots, p,$$
(1.1)

where β_j , $j = 0, \dots, p$, are unknown regression parameters and $\mathbf{x}_i = [1, x_{i1}, \dots, x_{ip}]'$ is the i th vector of fixed covariates.

The estimation of the logit model parameters results in having a statistical relation between θ_i and the corresponding i th vector of fixed covariates. This comes by modelling the binomial distribution parameters (θ_i 's) as a cumulative density function (cdf) of the logistic distribution. The logistic distribution has the following probability density function (pdf),

$$f(x) = \frac{e^x}{(1 + e^x)^2} = \frac{e^{-x}}{(1 + e^{-x})^2}, \quad -\infty < x < \infty, \quad (1.2)$$

and a cdf as

$$F(t) = \int_{-\infty}^t \frac{e^x}{(1 + e^x)^2} dx = \int_{-\infty}^t \frac{e^{-x}}{(1 + e^{-x})^2} dx = \frac{e^t}{1 + e^t} = \frac{1}{1 + e^{-t}}, \quad -\infty < t < \infty. \quad (1.3)$$

The logistic cdf given in (1.3) matches with (1.1) where t is expressed as $t = \mathbf{x}'\boldsymbol{\beta}$. Figure 1.1 presents logistic cdf's for different values of regression parameters β_0 and β_1 . Figure 1.1 shows four plots where the top ones are for increasing logistic functions with $\beta_1 > 0$, but the lower ones are for decreasing logistic functions with $\beta_1 < 0$. The top-left plot fixes β_0 and changes β_1 which impacts the sharpness of logistic curve S-shape. The top-right plot fixes β_1 and changes β_0 which results in shifting the logistic curve either to the left or right. A similar thing shows in bottom plots for decreasing logistic functions.

The application of the logit model generally aims to study the effects of the fixed covariates on the predicted probabilities of occurrence. Therefore, there is a focus on estimates and confidence intervals of the regression parameters.

This thesis will investigate applying the imprecise probabilistic methods to the logit model under the Bayesian approach. This requires introducing imprecise probabilities and their models as a main goal of this chapter. Understanding probabilistic imprecision comes after introducing the Bayesian robust and hierarchical methods with a general look at historical developments of Bayesian methods.

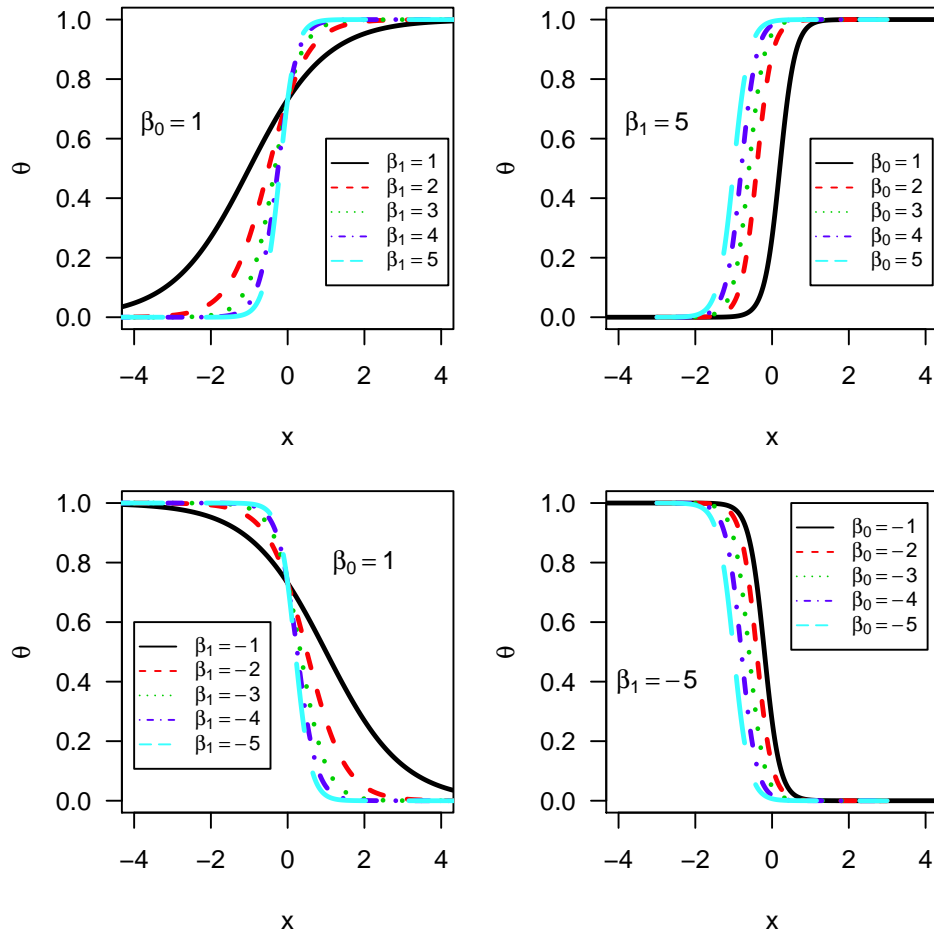


Figure 1.1: Plots of different logistic cdf's.

This introductory chapter is organized as follows. Section 1.2 starts by introducing some technicalities of the logit model. Mathematical aspects are given to describe the relation between the logit model and the exponential families. Also, an example of a real life application of the logit model is given in order to show its importance to the world of applied statistics.

Section 1.3 explains how the Bayesian approach is applied to the logit model. This requires introducing prior and posterior distributions of the main regression parameters in the logit model.

Bayesian robustness in the logit model is discussed in Section 1.4. Understanding robust methods helps to visualize and handle the concept of probabilistic imprecision with less difficulty. Furthermore, more details are given about the history of using and preferring robust methods in the Bayesian approach. A mathematical description of robust methods is given for building interval estimates for parameters of interest.

Section 1.5 takes the reader into a quick historical trip of Bayesian methods developments and applications. Basic and foundational studies in Bayesian methods will be cited. The main goal is to know circumstances that led probabilists, statisticians and researchers to adopt the imprecise probabilistic approach.

Section 1.6 goes deeper and extends sections 1.4 and 1.5 by presenting imprecise probabilities. Section 1.6 will prepare to jump smoothly to Section 1.7 in which the imprecise Dirichlet model is introduced.

In Section 1.7, two imprecise probability models are given: Imprecise Beta Model (IBM) as given in Walley (1991) and its generalized form, the Imprecise Dirichlet Model (IDM), as given in Walley (1996) and Bernard (2005). The study of IBM and IDM is necessary before going into two more imprecise probability models in Chapter 3, the imprecise logit-normal model (ILnM) and its generalized form, the imprecise multivariate logit-normal Model (IMLnM), where both of them are given in Bickis (2009). Aspects of probabilistic imprecision will be shown in IBM and IDM provided that Walley (1991), Walley (1996), Bernard (2005) and Bickis (2009) are considered as driving references and main sources of knowledge.

Section 1.8 gives examples of applications of imprecise probability models to show usefulness of thinking in probabilistic imprecision as an approach in real life problems.

Applications include several fields as classification, clinical trials and regression.

Finally, Section 1.9 looks at the general design of this thesis. A brief description of next chapters is given to prepare to move from theoretical forms of imprecise probability models to their new application in the logit model.

1.2 Logit Model

The logit model is a generalized linear model (GLM). In a GLM, a random variable Y follows a probability distribution that belongs to a one-dimensional exponential family. The distribution of a random variable is said to belong to a one-dimensional exponential family of distributions if its probability density function takes the form of

$$f(y|\eta, \phi) = \exp \left\{ \frac{g(y)\eta - b(\eta)}{a(\phi)} + c[g(y), \phi] \right\}, \quad (1.4)$$

where η is called the canonical parameter, a is an arbitrary function, b is differentiable, c is a function that lets $f(\cdot)$ to integrate to 1 and ϕ , being known, is the dispersion parameter. The binomial family is a well known exponential family of distributions. This is shown as follows

$$\begin{aligned} f(y|\eta, \phi) &= \binom{n}{y} \theta^y (1 - \theta)^{n-y} \\ &= \exp \left\{ \log \left[\binom{n}{y} \theta^y (1 - \theta)^{n-y} \right] \right\} \\ &= \exp \left[y \log \theta + (n - y) \log(1 - \theta) + \log \binom{n}{y} \right] \\ &= \exp \left[y \log \left(\frac{\theta}{1 - \theta} \right) + n \log(1 - \theta) + \log \binom{n}{y} \right], \end{aligned} \quad (1.5)$$

where $f(\cdot)$ is the density function (with respect to a counting measure) and

$$\begin{aligned}
& \eta = \log \left(\frac{\theta}{1 - \theta} \right) \\
\Rightarrow & e^\eta = \frac{\theta}{1 - \theta} \\
\Rightarrow & \theta = \frac{e^\eta}{1 + e^\eta} \\
\Rightarrow & 1 - \theta = \frac{1}{1 + e^\eta}.
\end{aligned} \tag{1.6}$$

Then

$$b(\eta) = -n \log(1 - \theta) = n \log[1 + \exp(\eta)]. \tag{1.7}$$

Also,

$$c(y, \phi) = \log \left[\binom{n}{y} \right], \quad a(\phi) = 1 \quad \text{and} \quad g(y) = y. \tag{1.8}$$

If $g(y) = y$ in (1.4) which is the case in (1.5), then the random variable Y has the following mean and variance

$$E(Y) = \mu = b'(\eta), \tag{1.9}$$

$$V(Y) = b''(\eta) a(\phi). \tag{1.10}$$

Generalized linear models are discussed in detail in McCullagh & Nelder (1989) and McCullagh & Searle (2001).

If the logit model has two regression parameters, β_0 and β_1 , then the i th probability of occurrence θ_i is modelled as

$$\theta_i = \theta(x_i) = \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} = \frac{1}{1 + e^{-(\beta_0 + x_i \beta_1)}}, \quad 0 < \theta_i < 1, \quad i = 1, \dots, m, \tag{1.11}$$

which is re-written as

$$\log \left(\frac{\theta_i}{1 - \theta_i} \right) = \beta_0 + x_i \beta_1. \tag{1.12}$$

The left hand side of (1.12) is referred to as the logit function. It can be seen that θ_i in (1.11) has the following limits,

$$\lim_{x_i \rightarrow -\infty} \theta(x_i) = 0 \quad \text{and} \quad \lim_{x_i \rightarrow \infty} \theta(x_i) = 1, \quad \beta_1 > 0, \tag{1.13}$$

and

$$\lim_{x_i \rightarrow -\infty} \theta(x_i) = 1 \quad \text{and} \quad \lim_{x_i \rightarrow \infty} \theta(x_i) = 0, \quad \beta_1 < 0. \quad (1.14)$$

Limits in (1.13) are identical to the limits of the logistic cdf given in (1.3) when $\beta_0 = 0$ and $\beta_1 = 1$.

The logit model describes various real life phenomena. For example, Berkson (1944) made an application of logit model in biological assay to find statistical relations between the amount of dose and response (deaths). This relation is called in bioassay as the dose-response relationship. It can be built by fitting a logit model, say with two parameters β_0 and β_1 , such that the predicted responses correspond to given doses. Having a dose-response relation gives an opportunity to make comparisons among drugs in terms of potency. In bioassay experiments, different doses are given to groups of experimental individuals or animals. The number of responses in each group to a given dose is considered as an observation on a binomial random variable. In this type of experiment, a parameter θ defines the tolerance of the animal on a specific dose x with no response.

In bioassay, there is often an interest in the dose amount that produces a response in 50% of the experimented animals. This dose amount is called the median effective dose and is referred to as the ED_{50} value. If the response proportion is 90%, then the effective dose is the ED_{90} value, and so on. Subject to (1.12), the dose for which $\theta = 0.5$ is given as

$$\log \left(\frac{0.5}{1 - 0.5} \right) = 0 = \beta_0 + \beta_1 (ED_{50}), \quad (1.15)$$

which means that

$$ED_{50} = \frac{-\beta_0}{\beta_1}. \quad (1.16)$$

Once the logit model is fitted, the parameters in (1.16) are replaced by the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$.

Estimation of regression parameters requires following a statistical approach. This includes the maximum likelihood and the Bayesian approaches. The Bayesian approach will be presented in Section 1.3.

1.3 Bayesian Approach for the Logit Model

The Bayesian approach states that if Y is a random variable with a probability density function $f(y|\theta)$, and θ is a distribution parameter that follows a prior distribution $\pi(\theta)$, then the posterior distribution $\pi(\theta|y)$ according to Bayes rule is given by

$$\pi(\theta|y) = \frac{(\text{likelihood})(\text{prior})}{\text{marginal}} = \frac{f(y|\theta) \pi(\theta)}{m(y)}, \quad (1.17)$$

where $m(y)$ is the marginal distribution of y defined as

$$m(y) = \int_{\Theta} f(y|\theta) \pi(\theta) d\theta = \int_{\Theta} f(y, \theta) d\theta. \quad (1.18)$$

The Bayesian approach allows representation of pre-experimental beliefs on θ . This requires to assign a prior distribution in order to produce the posterior distribution $\pi(\theta|y)$ that is conditioned on observed data. More details about Bayesian methods are in Bernardo and Smith (1994) and Robert (2001) and information about the history of Bayesian methods is given in Stigler (1983, 1986).

In Bayesian analysis of the logit model, a prior distribution is assigned to regression parameters $\beta = [\beta_0, \dots, \beta_p]'$. The regression parameter vector is considered as a random one so that it follows a joint probability distribution called the joint prior distribution, denoted by

$$\beta \sim \pi(\beta). \quad (1.19)$$

The posterior distribution of β given observed values y_1, \dots, y_m , assuming independence among y_1, \dots, y_m , is determined as

$$\begin{aligned} \pi(\beta|\mathbf{y}) &= \pi(\beta|y_1, \dots, y_m) = \frac{f(\mathbf{y}, \beta)}{m(\mathbf{y})} = \frac{f(\mathbf{y}|\beta) \pi(\beta)}{m(y_1, \dots, y_m)} \\ &= \frac{\left[\prod_{i=1}^m f(y_i|\beta) \right] \cdot \pi(\beta)}{m(y_1, \dots, y_m)} = \frac{\left[\prod_{i=1}^m f(y_i|\beta) \right] \cdot \pi(\beta)}{\int_{-\infty}^{\infty} \left\{ \left[\prod_{i=1}^m f(y_i|\beta) \right] \pi(\beta) \right\} d(\beta)}, \end{aligned} \quad (1.20)$$

where $f(y_i|\beta)$ is the binomial distribution pdf, such that θ_i is modelled as in (1.1), and $m(\cdot)$ is the marginal joint pdf of \mathbf{y} . The posterior multivariate pdf is usually intractable,

so computational methods are implemented to infer $\pi(\boldsymbol{\beta}|\mathbf{y})$. To estimate $\boldsymbol{\beta}$, the posterior distribution is simulated, then $\boldsymbol{\beta}$ is averaged in (1.20).

Having a single prior π is not always the case, but a set of priors can be assumed. That is, a class of priors is given to the parameters of interest such that $\pi \in \mathcal{A}$, where \mathcal{A} refers to a set of prior distributions. This allows one to study robustness of Bayesian models to changes in the prior knowledge of statistical models parameters. Section 1.4 sheds light on Bayesian models robustness and reasons behind considering it.

1.4 Bayesian Models Robustness

Let $\mathcal{A}_{\text{prior}}$ be a set of priors on a single parameter θ , then there is a corresponding set of posteriors $\mathcal{A}_{\text{posterior}}$. If θ , in each single prior, is estimated by finding $E(\theta|y)$, analytically or computationally, then a set of Bayesian estimates is produced for θ . The robust Bayesian approach focuses on the range of the posterior means

$$\left[\inf_{\pi \in \mathcal{A}_{\text{prior}}} E(\theta|y), \sup_{\pi \in \mathcal{A}_{\text{prior}}} E(\theta|y) \right]. \quad (1.21)$$

The goal is to find the range of posterior estimates and conditions for shortening the interval estimate in (1.21). Gustafson (1996a) mentions that robust analysis of (1.21) is referred to as formal sensitivity analysis.

Regression parameters $\boldsymbol{\beta}$ in the logit model can have a set of prior probability distributions rather than being restricted to follow a precise one. Once the regression parameters are estimated from the produced posterior distributions, then interval estimates for all regression parameters $\boldsymbol{\beta}$ are made. This enables to study the effects of having prior variation on the interval estimate length for each regression parameter β_j , $j = 0, \dots, p$.

The interval estimate for each β_j , $j = 0, \dots, p$ is denoted by $\hat{\beta}_j \in [\underline{\hat{\beta}}_j, \bar{\hat{\beta}}_j]$ where $\underline{\hat{\beta}}_j$ refers to the lower estimate and $\bar{\hat{\beta}}_j$ is the upper estimate. In this case, there is

$$\left[\inf_{\pi \in \mathcal{A}_{\text{prior}}} E(\boldsymbol{\beta}|y_1, \dots, y_m), \sup_{\pi \in \mathcal{A}_{\text{prior}}} E(\boldsymbol{\beta}|y_1, \dots, y_m) \right], \quad (1.22)$$

and

$$\left[\hat{\underline{\theta}} = \inf_{\pi \in \mathcal{A}_{\text{prior}}} E(\theta|y_1, \dots, y_m), \bar{\hat{\theta}} = \sup_{\pi \in \mathcal{A}_{\text{prior}}} E(\theta|y_1, \dots, y_m) \right]. \quad (1.23)$$

The idea of prior changes have caught attention in past through various applications. Greenhouse and Wasserman (1995) computed ranges of posterior expectations under priors in ϵ -contaminated class of priors. The ϵ -contaminated class of priors contains all mixture densities of the form

$$(1 - \epsilon) p(\theta) + \epsilon q(\theta), \quad (1.24)$$

where q can be any density. A useful source of knowledge to navigate methods and techniques in robust Bayesian modelling is Berger (1985).

Robustness can be applied in hierarchical models as in Gustafson (1996b). Hierarchical models are given in Lehmann and Casella (1998). The following simple example gives an idea about hierarchical Bayesian modelling. Let $X \sim N(\theta, c)$, where c is a known constant, and let the prior distribution of θ be $N(\mu, \tau^2)$ given that μ is known. Then τ^2 is assumed to follow a distribution, say $\tau^2 \sim \text{gamma}(\alpha, \beta)$, given that the hyperparameters α and β are known.

Robust methods have applications in testing hypotheses. Having a set of priors results in finding a lower bound for posterior probability of a null hypothesis. Sets of priors are considered in hypothesis testing by various authors. Berger and Delampady (1987) and Berger and Sellke (1987) demonstrated that a set of priors may produce a lower bound of the posterior probability of a true null hypothesis that is larger than the corresponding p -value. Carlin and Sargent (1996) tried different prior assumptions in hypothesis testing for a parameter θ that is assumed to change over a determined interval.

The involvement of robust methods in clinical trials has reserved a place in applied statistics. Robust methods for clinical trials are used in Greenhouse and Wasserman (1995). Spiegelhalter, Freedman and Parmar (1994) suggested three different priors in clinical trials experiments for a certain parameter of interest. If the clinical trials experiments cause a concern of beliefs about priors in the proposed model, then the priors are considered robust if the posterior distributions make similar results.

Bayesian methods literature includes useful review studies of robust methods. Berger (1990, 1994) and Wasserman (1992) are good examples that treat the robustness of Bayesian estimates to prior distributions changes. Pericchi and Walley (1991) is one more advanced source of robustness methods but for Bayesian credible intervals. Walley (1991) is an excellent source of knowledge that discusses in depth the use of classes of priors.

Walley specifies classes of priors called imprecise probability models. Walley (1991) is considered as a very standard and rigorous work. Walley (1991) has to be cited in any work on imprecise probability models and assessments of Bayesian robustness. Walley's foundational book is considered by Berger (1994) as : *“this latter work is particularly to be recommended for its deep and scholarly study of the foundations of imprecision and robustness”*.

It is important to look at the historical developments of Bayesian methods that caused statisticians to consider and apply robust methods. Therefore, Section 1.5 goes into a historical trip of developments of Bayesian methods to have a deeper understanding of the applications of Bayesian approach to logit model. Section 1.5 attempts to form a bridge between previous and next sections of this chapter.

1.5 Historical Development of Bayesian Methods

Presenting the history of Bayesian methods development acknowledges reasons and circumstances that led statisticians to think of interval values as parameters estimates. The statistical literature provides several sources of knowledge about the historical establishment of Bayesian methods. Powerful studies are in Stigler (1986), Fienberg (2006) and Aldrich (2008).

Fienberg (2006) made a strongly recommended and well written work. Fienberg's professional work is the most considered one in this section. He presented rich and rigorous study of historical debates between Bayesian and non-Bayesian statisticians and probabilists.

This section is divided into three subsections. Subsection 1.5.1 starts by defining the principle of inverse probability in order to let the reader to see how mathematicians and probabilists in 19th century thought of probability conditional on observed data. Then Subsection 1.5.2 moves to the circumstances in which the frequentist approach appeared. Subsection 1.5.3 looks at research work in the 20th century regarding the establishment of the subjective Bayesian approach.

1.5.1 Principle of Inverse Probability

Early applications of Bayesian methods took into account the discrete uniform priors with Laplace’s principle of insufficient reason. The principle of insufficient reason implies that if n possibilities are only distinguishable by their names, then each possibility can be given a probability (prior) equal to $\frac{1}{n}$. Examples that apply this principle is in assigning probabilities to coins, dice, and cards. If a coin is thrown, then the results are either a “tail” or “head” and the probabilities of both is equal to $\frac{1}{2}$. Probabilistic equity applies to a symmetric die in which the probability of having any of the die faces after being thrown is $\frac{1}{6}$.

Fienberg (2006) mentions that the principle of insufficient reason was basically established by Laplace (1774) and took its general form in Laplace (1812). Laplace used the principle of insufficient reason to formulate the use of the uniform prior for the binomial distribution parameter. De Morgan (1837) renamed this principle as “principle of inverse probability”. De Morgan included the word “inverse” in the new name to infer backwards from data to probability.

The principle of inverse probability turned out to be a standard method of choice for the 19th century scientists. For example, Edgeworth (1883) made an astonishing derivation of Student’s t -distribution by establishing the posterior distribution of the mean μ of the normal distribution. Some details about the technicalities of Edgeworth contribution are given in Stigler (1978). Pearson (1907) discussed the need and importance of past experience to future expectation.

The principle of inverse probability was renamed by Keynes (1921) as the “principle of indifference”. This is because Keynes thought that the principle of inverse probability only applies to experiments with equal prior probabilities. More details about the history of using and applying the principle of inverse probability is in Dale (1999).

The principle of inverse probability was not always considered. Ronald Fisher, a British statistician, rejected it strongly which led to the birth of the frequentist approach. Subsection 1.5.2 moves to talk generally about the frequentist approach.

1.5.2 Appearance of the Frequentist Approach

Fisher (1922) established his own approach of inference that was called the likelihood approach. In fact, Fisher's approach was revolutionary to the statistical thinking where he introduced the likelihood approach and its use to find the maximum likelihood estimates. He went farther by presenting the concepts of sufficiency and efficiency. Aldrich (2008) stated what Fisher (1925) mentioned about the principle of inverse probability as : *"the theory of inverse probability is founded upon an error, and must be wholly rejected"*. Aldrich mentions that Fisher thought the Bayesian approach formulates the problem and produces a solution and then withholds it. Fisher developed his new approach in Fisher (1925) where he made the formal principles of tests of significance. The important contributions made by Fisher in the third decade of the 20th century were followed by a foundational axiomatic work in probability by Kolmogorov (1933, 1950).

The appearance of the frequentist approach did not kill the principle of indifference. The main struggle between supporters of the principle of indifference and the frequentist approach started during the first half of the 20th century. Keynes (1921) attempted to treat the concept of probability in a new way by allowing for the possibility of having a personal degree of belief. Keynes developed a new meaning of probability itself. The introduction of the new probabilistic concept led Keynes to have a great influence on the traditional understanding of probability. Keynes prepared for the rise of what will be called later the "subjective probability".

The allowance for degree of belief in probability was the seed to establish the subjective Bayesian approach. Subsection 1.5.3 gives more information about this approach and its growth.

1.5.3 Establishment of the Subjective Bayesian Approach

The study of developments and debates of Bayesian methods ends in touching the meaning of the subjective Bayesian approach. Goldstein (2006) is a recent study that pays attention to subjective approach as principle and practice. Machina and Schmeidler (1992) explored past contributions in subjective approach.

The subjective probability approach started to have its own popularity after Keynes

(1921). Ramsey (1926) thought that knowledge about probabilities is personal, therefore it is subjective.

De Finetti (1937) justified the concept of subjective probability by introducing the concept of exchangeability and the implicit role of prior distributions. Savage (1954) considered Keynes (1921) new meanings of probability as the earliest account of the modern concept of personal probability. This approves the strong influence of Keynes on Bayesian subjectivists.

The fifties of the 20th century was an important decade to subjective probability approach and subjectivists where more attention and publications arose to adopt, apply and promote Bayesian subjectivism. This was a normal result of the pioneering work of Savage (1954). Savage considered de Finetti (1937) notions to develop a non-frequentist approach. Savage foundational work established the birth of an advanced methodology of subjective methods. Savage (1962) completed his beliefs by introducing the subjective approach as a part of the foundations of statistical inference. Good (1976) thinks that Savage revived the understanding of the whole Bayesian methods.

The fast growth of the subjective Bayesian approach was going in parallel with other attempts of applying prior beliefs in statistical applications. For example, Fisher (1956) defined the Bayesian argument as a fiducial probability by prioritizing a single parameter θ in one – dimensional random variables with a uniform distribution. Fisher fiducial argument was considered later in Lindley (1958), but Fisher could not convince all researchers of his time to adopt his method. This is because Fisher looked forward to build a posterior degree of belief without mentioning a prior one. Fisher was critiqued by Savage (1961) as “*a bold attempt to make the Bayesian omelet without breaking the Bayesian eggs*”. Of course, Savage did not mean to underestimate Fisher himself, but he meant that the fiducial approach is used without a complete dependence and specification of a prior distribution.

The strength and richness of research work in subjectivism made the motivation to start developing and establishing the robust Bayesian methods. An early interest in robustness is in Box and Tiao (1961) with a focus on choosing suitable prior distributions. Edwards, Lindman and Savage (1963) applied robustness of statistical modelling in psychological research.

The establishment of subjectivism, being followed by starting studies of robust Bayesian methods, prepared to the birth of hierarchical Bayesian modelling. The allowance of the personal beliefs in the prior assumptions created a thriving interest of applying hierarchical Bayesian methods. Tiao and Tan (1965) presented a hierarchical normal model.

It is interesting to mention that Bayesian subjectivism was not the only approach to follow in Bayesian statistics. Jeffreys (1939) used the principle of indifference to update the degree of personal belief of probabilities and to derive what he called the objective priors.

In a Bayesian model, personal beliefs can be expressed by assigning a group of priors to parameters of interest. As mentioned in Section 1.4, if a class of priors is selected, then an interval estimate for a parameter of interest can be found. Therefore, assignment of priors class is itself a point of interest. How large should the class be so that the interval estimate is kept as short as possible? Imprecise probability theory can answer this important question by making use of the imprecise probability models. Section 1.6 presents basic concepts of imprecise probabilities depending on Walley (1991).

1.6 Imprecise Probabilities

In probability theory, an event A has a precise probability between 0 and 1 to quantify the chance of occurrence. An imprecise probability is thought of in case of having a vagueness or confliction in information for assessing the precise probability. Therefore, generalizing precise probabilities is important in order to have a general assessment of the event A . This requires replacing the probability point value by an interval one.

An imprecise probability for an event A is made by assigning lower and upper probabilities. The imprecise probabilities literature denotes the lower probability by $\underline{P}(A)$ and the upper one by $\overline{P}(A)$. A precise probability comes from having equal lower and upper probabilities in an imprecise probability. The idea of an interval probability is not new and goes back long time ago. According to Jane Hutton in the discussion of Walley (1996), Ostrogradsky (1838) was the first to make an explicit use of lower and upper probabilities in the context of judicial decisions.

To illustrate the use of imprecise probabilities, the following example is taken from

Walley (1991). Let A denote an event that a particular thumbtack lands pin-up on its next toss. To construct lower and upper probabilities of A , several assessment strategies can be followed to use a relevant evidence about A . Evidences include records of previous tosses of this thumbtack or records from other thumbtacks. A simple strategy suggests to make intuitive assessments of $\underline{P}(A)$ and $\overline{P}(A)$. For example, if a thumbtack lands pin-up 3 times in 10 tosses, then an intuitive assessment makes low values for lower and upper probabilities of A , say $\underline{P}(A) = 0.1$ and $\overline{P}(A) = 0.4$. If the same thumbtack lands pin-up in 8 times, then lower and upper probabilities can be intuitively assessed by, say $\underline{P}(A) = 0.7$ and $\overline{P}(A) = 0.9$. Intuitive assessment is not a unique strategy. The lower and upper probabilities can be calculated based on observed values of a random variable which will be shown in Section 1.7.

Imprecise probabilities have their own applications in real world problems. Imprecise probabilities found their applications to the world of artificial intelligence early. Dempster (1967) used probabilistic imprecision to extend the belief functions. Belief functions were used in Shafer (1976). Theories in Shafer (1976) and Shafer and Vovk (2001) are used in De Cooman and Hermans (2008) in order to apply imprecise probabilities to trees with applications to artificial intelligence.

Other interesting applications are in applying imprecise probabilities to linguistics as in Zadeh (2002, 2006). This came after Zadeh (1975) where the concept of a linguistic variable with values as words or sentences was used in the artificial language to provide a basis for approximate reasoning. Imprecise probability theory can offer its services to the fields of finance and economics as in Vicig (2008). Imprecise probabilities are found to apply in engineering design by Aughenbaugh and Paredis (2006).

Imprecise probabilities stayed in shadows and depended on developments made by personal interests of mathematicians and statisticians until the nineties of the 20th century. The theory of imprecise probabilities gathered and received strong attention after Walley (1991). Walley presented a very rigid, foundational and strong theoretical description of imprecise probabilities. Walley went far in establishing the properties of imprecise probabilities and their uses in building the imprecise probability models. Walley refers to the interval probability by the term “imprecise probability”. However, the term “interval probability” was recently used in literature as in Weichselberger (2000).

Imprecise probabilities are meant to reflect the coherence of personal imprecise probabilistic assessments. Coherence is a self consistency requirement where an imprecise probability for an event is thought of as a result of thinking in how to handle a collection of probabilistic assessments.

To define lower and upper probabilities mathematically, let \mathcal{X} be a linear space of random variables on a sample space Ω , then a lower prevision, denoted by bolded $\underline{\mathbf{P}}$, is a real – valued function that maps \mathcal{X} to real numbers. That is

$$\underline{\mathbf{P}} : \mathcal{X} \longrightarrow \mathbb{R}. \quad (1.25)$$

A lower prevision $\underline{\mathbf{P}}$ is coherent when it is characterized by three axioms. The axioms are, for all $X \in \mathcal{X}$, $Y \in \mathcal{X}$ and positive λ ,

$$\left. \begin{aligned} \underline{\mathbf{P}}(X) &\geq \inf X, \\ \underline{\mathbf{P}}(\lambda X) &= \lambda \underline{\mathbf{P}}(X), \\ \underline{\mathbf{P}}(X + Y) &\geq \underline{\mathbf{P}}(X) + \underline{\mathbf{P}}(Y), \end{aligned} \right\} \quad (1.26)$$

where X is a real – valued function on Ω . Having $\underline{\mathbf{P}}$ and \mathcal{X} forms the lower prevision triple $(\Omega, \mathcal{X}, \underline{\mathbf{P}})$.

The coherent upper prevision is a conjugate to the lower one. That is, with an upper prevision triple $(\Omega, \mathcal{X}, \overline{\mathbf{P}})$, a coherent upper prevision, denoted by bolded $\overline{\mathbf{P}}$ and defined as

$$\overline{\mathbf{P}} : \mathcal{X} \longrightarrow \mathbb{R}, \quad (1.27)$$

is characterized by

$$\left. \begin{aligned} \overline{\mathbf{P}}(X) &\leq \sup X, \\ \overline{\mathbf{P}}(\lambda X) &= \lambda \overline{\mathbf{P}}(X), \\ \overline{\mathbf{P}}(X + Y) &\leq \overline{\mathbf{P}}(X) + \overline{\mathbf{P}}(Y). \end{aligned} \right\} \quad (1.28)$$

Both of $(\Omega, \mathcal{X}, \underline{\mathbf{P}})$ and $(\Omega, \mathcal{X}, \overline{\mathbf{P}})$ assume \mathcal{X} to be a linear space. The space \mathcal{X} is linear if for all $X \in \mathcal{X}$, $Y \in \mathcal{X}$ and $\lambda \in \mathbb{R}$, there is

$$\left. \begin{aligned} \lambda X &\in \mathcal{X}, \\ Z = X + Y &\longrightarrow Z \in \mathcal{X}. \end{aligned} \right\} \quad (1.29)$$

Based on (1.26) and (1.28), there is

$$\inf X \leq \underline{P}(X) \leq \overline{P}(X) \leq \sup X, \forall X \in \mathcal{X}. \quad (1.30)$$

Walley goes deeper by defining lower and upper probabilities for events. let \mathcal{A} be an arbitrary class of events, and consider it as a class of 0 – 1 (binary) random variables. Define the lower prevision \underline{P} on \mathcal{A} , then \underline{P} is called the lower probability on \mathcal{A} and $\underline{P}(A)$ is called the lower probability on an event $A \in \mathcal{A}$. Now, the upper probability comes by defining the upper prevision \overline{P} on \mathcal{A}^c , where $\mathcal{A}^c = \{A^c : A \in \mathcal{A}\} = \{1 - A : A \in \mathcal{A}\}$. This means that

$$\overline{P}(A) = 1 - \underline{P}(A^c). \quad (1.31)$$

Imprecise probabilities have certain properties. A basic property is that an imprecise probability for an event A can have a minimum and maximum bounds as

$$0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1, \quad (1.32)$$

where the degree of imprecision is defined as

$$\overline{P}(A) - \underline{P}(A). \quad (1.33)$$

Also

$$\underline{P}(\emptyset) = \overline{P}(\emptyset) = 0, \text{ and } \underline{P}(\Omega) = \overline{P}(\Omega) = 1. \quad (1.34)$$

Another property to know is that

$$\overline{P}\left(\bigcup_{j=1}^n A_j\right) \leq \sum_{j=1}^n \overline{P}(A_j). \quad (1.35)$$

If A_1, A_2, \dots are disjoint events then

$$\underline{P}\left(\bigcup_{j=1}^n A_j\right) \geq \sum_{j=1}^n \underline{P}(A_j), \quad (1.36)$$

and

$$\underline{P}\left(\bigcup_{j=1}^{\infty} A_j\right) \geq \sum_{j=1}^{\infty} \underline{P}(A_j). \quad (1.37)$$

Imprecise probabilities have their own statistical models. Walley (1991, 1996), Bernard (2005) and Bickis (2009) presented imprecise probability models with specific names. Section 1.7 looks at widely used imprecise probability model called the Imprecise Dirichlet Model.

1.7 Imprecise Dirichlet Model

An advanced look is given in this section on designing imprecise probabilities in clearly defined and marked models. Attention is paid to imprecise probability models used for the binomial distribution. This comes first by introducing the imprecise beta model as a special case of the imprecise Dirichlet model. The beta distribution is used frequently as a prior for the binomial distribution parameter, the probability of occurrence θ . If a class of beta priors is assigned to the binomial distribution parameter θ , then the posterior distributions will be expressed in closed forms. This is attractively usual since the beta distribution is a conjugate prior to the binomial distribution, that is the posterior distribution of θ is also a beta distribution with hyperparameters different from those in the beta prior. Then the concept of probabilistic imprecision becomes less difficult to handle and understand.

According to Casella and Berger (2002), if Y is a random variable such that $Y \sim \text{bin}(n, \theta)$ and $\theta \sim \text{beta}(\alpha, \beta)$, then the joint pdf of Y and θ is

$$\begin{aligned} f(y, \theta) &= f(y|\theta) \pi(\theta) = \left[\binom{n}{y} \theta^y (1-\theta)^{n-y} \right] \cdot \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \right] \\ &= \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{y+\alpha-1} (1-\theta)^{n-y+\beta-1}, \end{aligned} \quad (1.38)$$

where $y = 0, \dots, n$, $0 < \theta < 1$, $\alpha > 0$ and $\beta > 0$. The marginal pdf of Y is found by integrating (1.38) with respect to θ ,

$$m(y) = \int_0^1 f(y, \theta) d\theta = \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)}. \quad (1.39)$$

The probability density function in (1.39) is for a probability distribution known as the beta-binomial. Beta-binomial distribution is then used in Bayes rule to find the posterior distribution

$$\pi(\theta|y) = \frac{f(y, \theta)}{m(y)} = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)} \theta^{y+\alpha-1} (1-\theta)^{n-y+\beta-1}, \quad (1.40)$$

where the posterior pdf in (1.40) is for $\theta|y \sim \text{beta}(y+\alpha, n-y+\beta)$. Therefore, prior and posterior distributions of θ belong to the same family of beta distributions but with different parameters. Figure 1.2 shows a plot for both of beta prior and posterior pdf's.

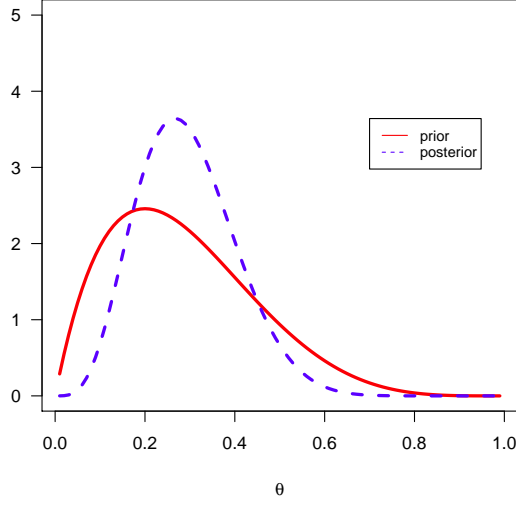


Figure 1.2: Plots of conjugate beta prior (solid) and posterior (dashed) pdf's of the binomial random variable parameter θ when $n = 10$, $y = 3$, $\alpha = 2$ and $\beta = 5$.

The Bayesian estimate (under the squared quadratic loss function) of the parameter θ is the mean of (1.40), that is

$$\hat{\theta} = E(\theta|y) = \int_0^1 \theta f(\theta|y) d\theta = \frac{y + \alpha}{n + \alpha + \beta}. \quad (1.41)$$

This simple Bayesian review is necessary to prepare to build the probabilistic imprecision skeleton in beta and Dirichlet distributions.

Now, a set of beta priors can be considered for the parameter θ . A set of beta distributions is particularly called the imprecise Beta model (IBM) by Walley (1991). Let $\theta \sim \text{beta}(\alpha_1, \alpha_2)$ (instead of $\text{beta}(\alpha, \beta)$) with the following alternative parametrization

$$\begin{aligned} \theta &\sim \text{beta}(\alpha_1, \alpha_2) = \text{beta}(\nu\varphi_1, \nu\varphi_2), \\ \text{where } \nu &= \alpha_1 + \alpha_2, \quad \varphi_1 = \frac{\alpha_1}{\alpha_1 + \alpha_2}, \quad \varphi_2 = \frac{\alpha_2}{\alpha_1 + \alpha_2}. \end{aligned} \quad (1.42)$$

Based on (1.40) and (1.42), the posterior distribution of θ is rewritten as

$$\theta|y \sim \text{beta}(y + \nu\varphi_1, n - y + \nu\varphi_2). \quad (1.43)$$

Note that for fixed $\nu > 0$, the set of priors is defined as

$$\mathcal{A}_{\text{prior}} = \{\text{beta}(\nu\varphi_1, \nu\varphi_2) : (\varphi_1, \varphi_2) \in \boldsymbol{\varphi}\}, \quad (1.44)$$

where φ is the parameter space of all possible values of φ_1 and φ_2 .

If IBM is considered, then there is a set of corresponding posteriors. The set of posteriors, given an observation y , is then defined as

$$\mathcal{A}_{\text{posterior}} = \{\text{beta}(y + \nu\varphi_1, n - y + \nu\varphi_2) : (\varphi_1, \varphi_2) \in \varphi\}. \quad (1.45)$$

Minimizing and maximizing $E(\theta)$ and $E(\theta|y)$ over the sets of prior and posterior distributions given in (1.44) and (1.45), with respect to $(\varphi_1, \varphi_2) \in \varphi$, gives lower and upper estimates in both sets of distributions. Walley (1991) refers to the lower and upper expectations by $\underline{E}(\theta)$ and $\overline{E}(\theta)$, $\underline{E}(\theta|y)$ and $\overline{E}(\theta|y)$.

Posterior lower and upper expectations come from estimating posterior distributions for which $\varphi_1 \rightarrow 0$ and $\varphi_1 \rightarrow 1$, respectively. If $\varphi_1 \rightarrow 0$ in (1.42) then $\varphi_2 \rightarrow 1$ and the posterior distribution in (1.43) becomes $\theta|y \sim \text{beta}(y, n - y + \nu)$, but if $\varphi_1 \rightarrow 1$ then $\varphi_2 \rightarrow 0$ and $\theta|y \sim \text{beta}(y + \nu, n - y)$. So, finding posterior expectations as in (1.41) for both of $\text{beta}(y, n - y + \nu)$ and $\text{beta}(y + \nu, n - y)$ gives the posterior lower and upper expectations of θ . For any θ in IBM, the prior lower and upper expectations are

$$\text{as } \varphi_1 \rightarrow 0 \quad \text{then} \quad \underline{E}(\theta) = 0, \quad (1.46)$$

$$\text{and} \quad \text{as } \varphi_1 \rightarrow 1 \quad \text{then} \quad \overline{E}(\theta) = 1, \quad (1.47)$$

while the posterior lower and upper expectations are

$$\text{as } \varphi_1 \rightarrow 0 \quad \text{then} \quad \underline{\hat{\theta}} = \underline{E}(\theta|y) = \frac{y}{n + \nu}, \quad (1.48)$$

$$\text{and} \quad \text{as } \varphi_1 \rightarrow 1 \quad \text{then} \quad \overline{\hat{\theta}} = \overline{E}(\theta|y) = \frac{y + \nu}{n + \nu}. \quad (1.49)$$

The lower and upper estimates in (1.48) and (1.49) are combined to form an imprecise estimate of θ as $\hat{\theta} \in [\underline{\hat{\theta}}, \overline{\hat{\theta}}]$. Figure 1.3 shows plots of lower and upper beta posterior cdf's when $\varphi_1 \rightarrow 0$ and $\varphi_1 \rightarrow 1$ for different binomial random variable values.

The lower and upper estimates in (1.48) and (1.49) can be applied to the thumbtack example given in Section 1.6. Fix $\nu = 1$, then for a thumbtack that lands pin-up 3 times in 10 tosses, the lower and upper estimates in (1.48) and (1.49) are

$$\left[\underline{\hat{\theta}} = \frac{3}{10 + 1} = \frac{3}{11}, \quad \overline{\hat{\theta}} = \frac{3 + 1}{10 + 1} = \frac{4}{11} \right]. \quad (1.50)$$

For a thumbtack that lands pin-up 8 times in 10 tosses, the lower and upper estimates are

$$\left[\hat{\theta} = \frac{8}{11} \quad , \quad \bar{\hat{\theta}} = \frac{9}{11} \right]. \quad (1.51)$$

It is important to distinguish the previous interval estimates from credible intervals in Bayesian methods. Recall that for a posterior distribution $P(\theta|y)$, then a 95% credible interval comes by finding c_1 and c_2 in

$$\int_{c_1}^{c_2} P(\theta|y) d\theta = 0.95, \quad (1.52)$$

where the credible interval $[c_1, c_2]$ is assigned with a probability of 0.95. On the contrary, the interval estimates in (1.50) and (1.51) are not assigned with probabilities but they come from finding the point estimates in an infinite family of posterior distributions as in (1.45). For each point estimate in an interval estimate, a corresponding credible interval can be constructed.

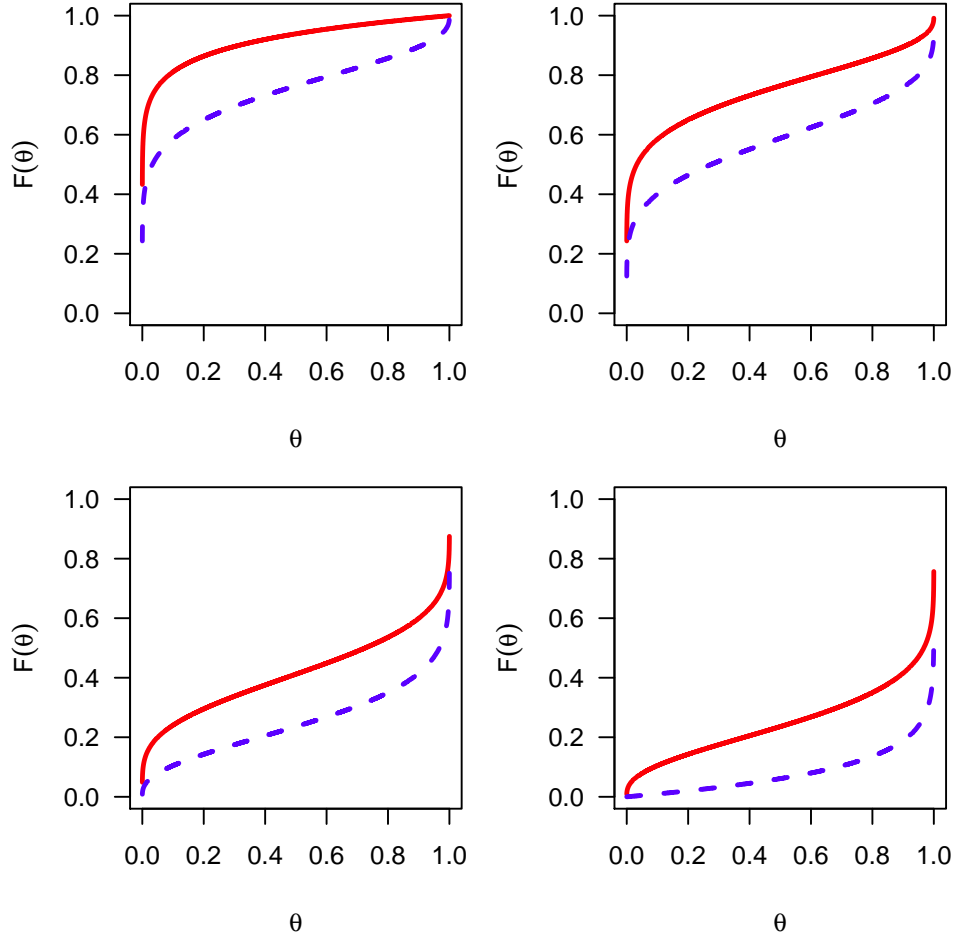


Figure 1.3: Plots of lower and upper beta posterior cdf's of θ when $\varphi_1 \rightarrow 0$ (solid) and $\varphi_1 \rightarrow 1$ (dashed) with $\nu = 2$, $n = 10$, $y = 1$ (top-left), $y = 3$ (top-right), $y = 7$ (bottom-left) and $y = 9$ (bottom-right).

The IBM is a special case of the Imprecise Dirichlet Model (IDM). The imprecise Dirichlet family of distributions forms a conjugate imprecise prior to the family of multinomial distributions. IDM was presented in Walley (1996) and discussed in Bernard (2005).

If \mathbf{Y} follows the multinomial distribution as

$$\mathbf{Y} = (Y_1, \dots, Y_m) \sim \text{mult}(N, \theta_1, \dots, \theta_m),$$

then \mathbf{Y} has the following pdf

$$\boxed{\begin{aligned} f(\mathbf{Y}|\boldsymbol{\theta}) &= \frac{N!}{m} \prod_{i=1}^m \theta_i^{y_i}, \\ \text{where } N &= \sum_{i=1}^m y_i, \sum_{i=1}^m \theta_i = 1, \quad 0 < \theta_i < 1, \quad \forall i. \end{aligned}} \quad (1.53)$$

If $\boldsymbol{\theta}$ follows the Dirichlet distribution as

$$\boldsymbol{\theta} = (\theta_1, \dots, \theta_m) \sim \text{Dir}(\alpha_1, \dots, \alpha_m), \quad (1.54)$$

then $\boldsymbol{\theta}$ has the following pdf

$$\boxed{\begin{aligned} f(\boldsymbol{\theta}|\boldsymbol{\alpha}) &= \frac{\Gamma\left(\sum_{i=1}^m \alpha_i\right)}{\prod_{i=1}^m \Gamma(\alpha_i)} \left(\prod_{i=1}^{m-1} \theta_i^{\alpha_i-1}\right) \left(1 - \sum_{i=1}^{m-1} \theta_i\right)^{\alpha_m-1}, \\ \text{where } \theta_m &\text{ is written as } \theta_m = 1 - \sum_{i=1}^{m-1} \theta_i, \quad \alpha_i > 0, \quad \forall i. \end{aligned}} \quad (1.55)$$

The posterior distribution of $\boldsymbol{\theta}|\mathbf{Y}$ is $\text{Dir}(y_1 + \alpha_1, \dots, y_m + \alpha_m)$ with pdf

$$\begin{aligned} f(\boldsymbol{\theta}|\boldsymbol{\alpha}) &= \frac{\Gamma\left(N + \sum_{i=1}^m \alpha_i\right)}{\prod_{i=1}^m \Gamma(y_i + \alpha_i)} \cdot \left(\prod_{i=1}^{m-1} \theta_i^{y_i + \alpha_i - 1}\right) \left(1 - \sum_{i=1}^{m-1} \theta_i\right)^{y_m + \alpha_m - 1}, \\ \text{where } \theta_m &\text{ is written as } \theta_m = 1 - \sum_{i=1}^{m-1} \theta_i, \quad \alpha_i > 0, \quad \forall i. \end{aligned} \quad (1.56)$$

Probabilistic imprecision in IDM is expressed in a similar way to IBM. Let

$$\nu = \sum_{i=1}^m \alpha_i = \alpha_1 + \dots + \alpha_m, \quad \varphi_i = \frac{\alpha_i}{\nu}, \quad \varphi_i \in [0, 1], \quad \forall i. \quad (1.57)$$

For any θ_i in IDM, the prior lower and upper expectations are

$$\text{as } \varphi_i \rightarrow 0 \quad \text{then} \quad \underline{E}(\theta_i) = 0, \quad (1.58)$$

$$\text{and} \quad \text{as } \varphi_i \rightarrow 1 \quad \text{then} \quad \overline{E}(\theta_i) = 1, \quad (1.59)$$

while the posterior lower and upper expectations are

$$\text{as } \varphi_i \rightarrow 0 \quad \text{then} \quad \hat{\theta}_i = \underline{E}(\theta_i | y_1, \dots, y_m) = \frac{y_i}{n + \sum_{i=1}^m \alpha_i} = \frac{y_i}{n + \nu}, \quad (1.60)$$

$$\text{and} \quad \text{as } \varphi_i \rightarrow 1 \quad \text{then} \quad \bar{\theta}_i = \overline{E}(\theta_i | y_1, \dots, y_m) = \frac{y_i + \sum_{i=1}^m \alpha_i}{n + \sum_{i=1}^m \alpha_i} = \frac{y_i + \nu}{n + \nu}. \quad (1.61)$$

Imprecise probability models have their own valuable applications. Section 1.8 gives examples of research fields in which imprecise probability models are found useful.

1.8 Applications of Imprecise Probability Models

Imprecise probability models have applications in several fields as clinical trials, classification and artificial intelligence. For example, Walley, Gurrin and Burton (1996) made a powerful study and application of the IBM to clinical trials analysis. IBM was applied for prior ignorance about binomial distribution parameter (probability of occurrence) to analyze clinical trials data. The imprecise probabilistic method was applied to data from clinical trials of extracorporeal membrane oxygenation (ECMO). The study mainly focused on the chance (probability) of survival for new-born babies who were treated by ECMO due to having acute respiratory failure or certain anomalies in cardiovascular circulation. An interesting application is in Coolen (1997) in which failure data are modelled imprecisely under IDM.

Imprecise probability models found their track to applications in statistical classification. Zaffalon (2002) made use of imprecise probabilities to create a new classifier called naive credal classifier (NCC). The imprecise classifier NCC extended a well known classifier called naive Bayes classifier (NBC). In NBC, if C is a class variable that takes values c_1, \dots, c_n , and A_1, \dots, A_n is a set of attributes variables where each attribute takes values a_{i1}, \dots, a_{ik} , then an instance of attributes, say a_1, \dots, a_n , is classified to a class $c_i \in C$ by maximizing

$$P(C|a_1, \dots, a_n) = \frac{P(a_1, \dots, a_n|C) P(C)}{P(a_1, \dots, a_n)} \propto P(a_1, \dots, a_n|C) P(C), \quad (1.62)$$

where independence is assumed among attributes, as in Duda and Hart (1973), conditional on the class variable

$$P(A_1, \dots, A_n|C) = \prod_{i=1}^n P(A_i|C). \quad (1.63)$$

Then (1.62) becomes

$$P(C|a_1, \dots, a_n) \propto P(C) \prod_{i=1}^n P(a_i|C). \quad (1.64)$$

The classifier NCC maps instances of attributes to a subset of classes in C rather than only a single class. NCC is used when $P(a_i|C)$ is interval-valued, $\forall i$ in (1.64).

Imprecise probability models were used in regression analysis by Walter, Augustin and Peters (2007), in which the imprecise normal model was considered as an imprecise prior to the parameters of linear regression model. This application followed a remarkable work done by Quaeghebeur and De Cooman (2005). Quaeghebeur and De Cooman built a general theory for using imprecise probability models in exponential families. Quaeghebeur and De Cooman (2005) came after a study of imprecise conjugate priors for one parameter exponential family made by Coolen (1993).

An interesting application of imprecise probabilities is in Cozman (2000) where Bayesian networks are extended to credal networks. Credal networks are directed acyclic graphs that are associated with sets of probability measures. Cozman (2005) established the theory of graphical models for imprecise probabilities, where “graphical models” include credal networks and other types of graphs as the undirected graphs.

The last few years had excellent developments in the theoretical aspects of imprecise probabilities. De Cooman, Hermans and Quaeghebeur (2009) applied the imprecise transition probabilities to Markov chains. This may pace the track for a revolutionary field of applications of Markov chains and to establish what could be later called as the Imprecise Markov Chains. Imprecise probabilities have passed expectations to have applications in decision theory in Troffaes (2007) and in minimum distance estimation in Hable (2010).

According to Augustin and Hable (2010), imprecise probability theory can generalize the results of robust statistics. The theory of imprecise probabilities is promising enough to present imprecise hierarchical Bayesian models as in De Cooman (2002).

1.9 Organization of the Thesis

This thesis is organized as follows. Chapter 2 gives a new method of inducing a prior distribution for regression parameters in logistic regression model.

Chapter 3 dives into the main body of this thesis where a new statistical regression model called ImpLogit is established, developed and fitted. ImpLogit is a new coined model that makes use of IDM and ILnM as imprecise priors for logistic regression parameters. ImpLogit model is presented as a promising and novel model. ImpLogit model is fitted in purpose of finding conditions in which imprecision (interval estimates lengths) is reduced in regression parameters estimates.

A detailed simulation study is given in Chapter 4. The simulation study aims to explore statistical behaviour aspects in 2-parameter ImpLogit model. The relation between the 2-parameter ImpLogit model design and the amount of imprecision in regression parameters is a point of main interest. It is important to see how imprecision of estimates of β_0 and β_1 , relates to covariate ranges and values of imprecise priors parameters.

The conclusion and future research plan on ImpLogit model are given in Chapter 5.

CHAPTER 2

BAYESIAN 2-PARAMETER LOGIT MODEL

This chapter goes into Bayesian modelling of the logit model with only two regression parameters, β_0 and β_1 . Cases of increasing and decreasing logistic curves are considered in sections 2.1 and 2.2, respectively. The Dirichlet distribution is considered as a prior distribution in logit model. Sections 2.1 and 2.2 show how to induce a prior joint pdf for $\boldsymbol{\beta} = [\beta_0, \beta_1]'$.

Section 2.3 introduces the logit-normal distribution. The multivariate form of logit-normal distribution is presented in Section 2.4. The prior distribution is shown to play as a conjugate one for the posterior distribution of the binomial distribution parameter θ .

2.1 Increasing 2-Parameter Logit Model

In the logit model, it is reasonable to have correlation among the binomial distribution parameters $\theta_i, \forall i$. Correlation comes by noticing that an increasing logistic curve means mathematically that $\theta_1 < \theta_2 < \dots < \theta_m$, for distinct x_i values. This implies having correlation between θ_{i-1} and θ_i since θ_{i-1} restricts the value of θ_i . The same thing is thought of in the decreasing logistic curve where $\theta_1 > \theta_2 > \dots > \theta_m$.

The prior distribution of the successive differences, $\theta_1, \theta_2 - \theta_1, \dots, \theta_m - \theta_{m-1}, 1 - \theta_m$, in the logistic curve is assumed to follow the Dirichlet distribution as

$$\begin{aligned} &(\theta_1, \theta_2 - \theta_1, \dots, \theta_m - \theta_{m-1}, 1 - \theta_m) \\ &\sim \text{Dir}(\alpha_1, \alpha_2 - \alpha_1, \dots, \alpha_m - \alpha_{m-1}, \nu - \alpha_m), \end{aligned} \quad (2.1)$$

such that $\theta_1 < \theta_2 < \dots < \theta_m$ and $\alpha_1 + \alpha_2 - \alpha_1 + \dots + \alpha_m - \alpha_{m-1} + \nu - \alpha_m = \nu$. Then the

hyperparameters in (2.1) are reparametrized as in (1.42)

$$\begin{aligned} &(\theta_1, \theta_2 - \theta_1, \dots, \theta_m - \theta_{m-1}, 1 - \theta_m) \\ &\sim \text{Dir}(\nu H_1, \nu H_2 - \nu H_1, \dots, \nu H_m - \nu H_{m-1}, \nu - \nu H_m), \end{aligned} \quad (2.2)$$

where $0 \leq H_1 < H_2 < \dots < H_m \leq 1$.

A prior joint distribution for $\boldsymbol{\beta} = [\beta_0, \beta_1]'$ can be built by assigning a joint distribution for two θ 's, say θ_j, θ_k . Such a joint prior with a pdf $f(\theta_j, \theta_k)$ is then used to induce a joint pdf $\pi(\beta_0, \beta_1)$ by applying the method of change of variables. To find $f(\theta_j, \theta_k)$, the aggregation property of the Dirichlet distribution is recalled. Briefly, the aggregation property is described as follows : if

$$\boldsymbol{\theta} = (\theta_1, \dots, \theta_m) \sim \text{Dir}(\alpha_1, \dots, \alpha_m), \quad (2.3)$$

then

$$\boldsymbol{\theta}^* = (\theta_1, \dots, \theta_j + \theta_k, \dots, \theta_m) \sim \text{Dir}(\alpha_1, \dots, \alpha_j + \alpha_k, \dots, \alpha_m), \quad \forall j, k, \quad (2.4)$$

and by applying this property to (2.2), the joint distribution of θ_j, θ_k is

$$(\theta_j, \theta_k - \theta_j, 1 - \theta_k) \sim \text{Dir}(\nu H_j, \nu H_k - \nu H_j, \nu - \nu H_k). \quad (2.5)$$

This implies

$$\theta_j \sim \text{beta}(\nu H_j, \nu - \nu H_j), \quad (2.6)$$

and

$$\theta_k \sim \text{beta}(\nu H_k, \nu - \nu H_k). \quad (2.7)$$

Based on (2.5) and (1.55), θ_j and θ_k have the following joint pdf (because the Jacobian determinant $|J| = 1$)

$$\begin{aligned} f(\theta_j, \theta_k) &= \frac{\Gamma(\nu)}{\Gamma(\nu H_j) \Gamma(\nu H_k - \nu H_j) \Gamma(\nu - \nu H_k)} \\ &\quad \times (\theta_j)^{\alpha H_j - 1} (\theta_k - \theta_j)^{\nu H_k - \nu H_j - 1} (1 - \theta_k)^{\nu - \nu H_k - 1}, \\ &\quad 0 < \theta_j < \theta_k < 1. \end{aligned} \quad (2.8)$$

Then (2.8) is written in terms of β_0 and β_1 by applying the change of variables method

$$\pi(\boldsymbol{\beta}) = f[\boldsymbol{\theta}(\boldsymbol{\beta})] \left| \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\beta}} \right|, \quad (2.9)$$

where the Jacobian matrix determinant is defined as

$$\left| \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\beta}} \right| = \begin{vmatrix} \frac{\partial \theta_j}{\partial \beta_0} & \frac{\partial \theta_j}{\partial \beta_1} \\ \frac{\partial \theta_k}{\partial \beta_0} & \frac{\partial \theta_k}{\partial \beta_1} \end{vmatrix} = \begin{vmatrix} \frac{e^{\beta_0+x_j\beta_1}}{(1+e^{\beta_0+x_j\beta_1})^2} & \frac{x_j e^{\beta_0+x_j\beta_1}}{(1+e^{\beta_0+x_j\beta_1})^2} \\ \frac{e^{\beta_0+x_k\beta_1}}{(1+e^{\beta_0+x_k\beta_1})^2} & \frac{x_k e^{\beta_0+x_k\beta_1}}{(1+e^{\beta_0+x_k\beta_1})^2} \end{vmatrix},$$

which becomes

$$\begin{aligned} & \left| (x_k - x_j) \left\{ \left[\frac{e^{\beta_0+x_j\beta_1}}{(1+e^{\beta_0+x_j\beta_1})^2} \right] \left[\frac{e^{\beta_0+x_k\beta_1}}{(1+e^{\beta_0+x_k\beta_1})^2} \right] \right\} \right| \\ &= (x_k - x_j) \left\{ \left[\frac{e^{\beta_0+x_j\beta_1}}{(1+e^{\beta_0+x_j\beta_1})^2} \right] \left[\frac{e^{\beta_0+x_k\beta_1}}{(1+e^{\beta_0+x_k\beta_1})^2} \right] \right\}. \end{aligned} \quad (2.10)$$

The prior joint pdf is induced by substituting (2.10) in (2.9) as

$$\begin{aligned} \pi(\boldsymbol{\beta}) &= \pi(\beta_0, \beta_1) = \left[\frac{\Gamma(\nu)}{\Gamma(\nu H_j) \Gamma(\nu H_k - \nu H_j) \Gamma(\nu - \nu H_k)} \right] \\ &\times \left\{ [\theta(\beta_0 + x_j \beta_1)]^{\nu H_j - 1} \right\} \\ &\times \left\{ [\theta(\beta_0 + x_k \beta_1) - \theta(\beta_0 + x_j \beta_1)]^{\nu H_k - \nu H_j - 1} [1 - \theta(\beta_0 + x_k \beta_1)]^{\nu - \nu H_k - 1} \right\} \\ &\times (x_k - x_j) \left\{ \left[\frac{e^{\beta_0+x_j\beta_1}}{(1+e^{\beta_0+x_j\beta_1})^2} \right] \left[\frac{e^{\beta_0+x_k\beta_1}}{(1+e^{\beta_0+x_k\beta_1})^2} \right] \right\} \\ &= \left[\frac{\Gamma(\nu)}{\Gamma(\nu H_j) \Gamma(\nu H_k - \nu H_j) \Gamma(\nu - \nu H_k)} \right] \\ &\times \left[\left(\frac{e^{\beta_0+x_j\beta_1}}{1+e^{\beta_0+x_j\beta_1}} \right)^{\nu H_j - 1} \left(\frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}} - \frac{e^{\beta_0+x_j\beta_1}}{1+e^{\beta_0+x_j\beta_1}} \right)^{\nu H_k - \nu H_j - 1} \right. \\ &\quad \left. \cdot \left(1 - \frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}} \right)^{\nu - \nu H_k - 1} \right] \\ &\times (x_k - x_j) \left\{ \left[\frac{e^{\beta_0+x_j\beta_1}}{(1+e^{\beta_0+x_j\beta_1})^2} \right] \left[\frac{e^{\beta_0+x_k\beta_1}}{(1+e^{\beta_0+x_k\beta_1})^2} \right] \right\} \end{aligned} \quad (2.11)$$

The posterior joint pdf $\pi(\beta_0, \beta_1 | y_1, \dots, y_m)$ is then found by plugging (2.11) in (1.20).

Since

$$y_i \sim \text{bin}(\theta_i, n_i) \quad \text{and} \quad \theta_i = \frac{e^{\beta_0+x_i\beta_1}}{1+e^{\beta_0+x_i\beta_1}}, \quad \forall i,$$

then

$$\begin{aligned}
\pi(\boldsymbol{\beta}|\mathbf{y}) &= \pi(\beta_0, \beta_1|y_1, \dots, y_m) \propto f(\mathbf{y}, \boldsymbol{\beta}) \\
&= \left\{ \prod_{i=1}^m \left[\binom{n_i}{y_i} \left(\frac{e^{\beta_0+x_i\beta_1}}{1+e^{\beta_0+x_i\beta_1}} \right)^{y_i} \left(\frac{1}{1+e^{\beta_0+x_i\beta_1}} \right)^{n_i-y_i} \right] \right\} \\
&\quad \times \left[\frac{\Gamma(\nu)}{\Gamma(\nu H_j) \Gamma(\nu H_k - \nu H_j) \Gamma(\nu - \nu H_k)} \right] \\
&\quad \times \left[\left(\frac{e^{\beta_0+x_j\beta_1}}{1+e^{\beta_0+x_j\beta_1}} \right)^{\nu H_j-1} \left(\frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}} - \frac{e^{\beta_0+x_j\beta_1}}{1+e^{\beta_0+x_j\beta_1}} \right)^{\nu H_k-\nu H_j-1} \right. \\
&\quad \left. \cdot \left(1 - \frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}} \right)^{\nu-\nu H_k-1} \right] \\
&\quad \times (x_k - x_j) \left\{ \left[\frac{e^{\beta_0+x_j\beta_1}}{(1+e^{\beta_0+x_j\beta_1})^2} \right] \left[\frac{e^{\beta_0+x_k\beta_1}}{(1+e^{\beta_0+x_k\beta_1})^2} \right] \right\} \tag{2.12}
\end{aligned}$$

$$\begin{aligned}
&= \left\{ \prod_{i=1}^m \left[\binom{n_i}{y_i} \left(\frac{e^{\beta_0+x_i\beta_1}}{1+e^{\beta_0+x_i\beta_1}} \right)^{y_i} \left(\frac{1}{1+e^{\beta_0+x_i\beta_1}} \right)^{n_i-y_i} \right] \right\} \\
&\quad \times \left[\frac{\Gamma(\nu)}{\Gamma(\nu H_j) \Gamma(\nu H_k - \nu H_j) \Gamma(\nu - \nu H_k)} \right] \\
&\quad \times \left[\left(\frac{e^{\beta_0+x_j\beta_1}}{1+e^{\beta_0+x_j\beta_1}} \right)^{\nu H_j} \left(\frac{e^{\beta_0+x_k\beta_1} - e^{\beta_0+x_j\beta_1}}{(1+e^{\beta_0+x_k\beta_1})(1+e^{\beta_0+x_j\beta_1})} \right)^{\nu H_k-\nu H_j} \right. \\
&\quad \left. \cdot \left(\frac{1}{1+e^{\beta_0+x_k\beta_1}} \right)^{\nu-\nu H_k} \right] \\
&\quad \times \left(\frac{1+e^{\beta_0+x_j\beta_1}}{e^{\beta_0+x_j\beta_1}} \right) \cdot \left(\frac{(1+e^{\beta_0+x_k\beta_1})(1+e^{\beta_0+x_j\beta_1})}{e^{\beta_0+x_k\beta_1} - e^{\beta_0+x_j\beta_1}} \right) \cdot (1+e^{\beta_0+x_k\beta_1}) \\
&\quad \times (x_k - x_j) \left\{ \left[\frac{e^{\beta_0+x_j\beta_1}}{(1+e^{\beta_0+x_j\beta_1})^2} \right] \left[\frac{e^{\beta_0+x_k\beta_1}}{(1+e^{\beta_0+x_k\beta_1})^2} \right] \right\} \\
&= \left\{ \prod_{i=1}^m \left[\binom{n_i}{y_i} \left(\frac{e^{\beta_0+x_i\beta_1}}{1+e^{\beta_0+x_i\beta_1}} \right)^{y_i} \left(\frac{1}{1+e^{\beta_0+x_i\beta_1}} \right)^{n_i-y_i} \right] \right\} \\
&\quad \times \left[\frac{\Gamma(\nu)}{\Gamma(\nu H_j) \Gamma(\nu H_k - \nu H_j) \Gamma(\nu - \nu H_k)} \right] \\
&\quad \times \left[\left(\frac{e^{\beta_0+x_j\beta_1}}{1+e^{\beta_0+x_j\beta_1}} \right)^{\nu H_j} \left(\frac{e^{\beta_0+x_k\beta_1} - e^{\beta_0+x_j\beta_1}}{(1+e^{\beta_0+x_k\beta_1})(1+e^{\beta_0+x_j\beta_1})} \right)^{\nu H_k-\nu H_j} \right. \\
&\quad \left. \cdot \left(\frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}} \right)^{\nu-\nu H_k} \right] \cdot \left[(x_k - x_j) \left(\frac{e^{\beta_0+x_k\beta_1}}{e^{\beta_0+x_k\beta_1} - e^{\beta_0+x_j\beta_1}} \right) \right] \cdot \tag{2.13}
\end{aligned}$$

The assignment of joint prior distribution in (2.5) can be made for any j, k , where $\theta_j < \theta_k$. Therefore, it becomes important to find the prior correlation between any selected θ_j and θ_k . For Dirichlet distribution, if

$$\boldsymbol{\theta} = (\theta_1, \dots, \theta_m) \sim \text{Dir}(\alpha_1, \dots, \alpha_m), \quad (2.14)$$

then according to Kotz, Balakrishnan and Johnson (2000),

$$E(\theta_j) = \frac{\alpha_j}{\sum_{i=1}^m \alpha_i}, \quad (2.15)$$

$$\text{Var}(\theta_j) = \frac{\alpha_j \left(\sum_{i=1}^m \alpha_i - \alpha_j \right)}{\left(\sum_{i=1}^m \alpha_i \right)^2 \left(\sum_{i=1}^m \alpha_i + 1 \right)}, \quad (2.16)$$

$$\text{Cov}(\theta_j, \theta_k) = \frac{-\alpha_j \alpha_k}{\left(\sum_{i=1}^m \alpha_i \right)^2 \left(\sum_{i=1}^m \alpha_i + 1 \right)}. \quad (2.17)$$

To find the correlation between any θ_j and θ_k in (2.5), using (2.16) and (2.17),

$$\text{Cov}(\theta_j, \theta_k - \theta_j) = \text{Cov}(\theta_j, \theta_k) - \text{Var}(\theta_j) \quad (2.18)$$

\Rightarrow

$$\frac{-\nu H_j (\nu H_k - \nu H_j)}{\nu^2 (\nu + 1)} = \text{Cov}(\theta_j, \theta_k) - \frac{\nu H_j (\nu - \nu H_j)}{\nu^2 (\nu + 1)} \quad (2.19)$$

\Rightarrow

$$\begin{aligned} \text{Cov}(\theta_j, \theta_k) &= \frac{\nu^2 H_j - (\nu H_j)^2 - \nu^2 H_j H_k + (\nu H_j)^2}{\nu^2 (\nu + 1)} = \frac{\nu^2 H_j (1 - H_k)}{\nu^2 (\nu + 1)} \\ &= \frac{H_j (1 - H_k)}{\nu + 1}. \end{aligned} \quad (2.20)$$

The covariance between any prior parameters θ_j and θ_k in (2.20) is always positive since having $\theta_j < \theta_k$, $\forall j, k$, requires θ_k to be positively correlated to θ_j . This is different from the covariance in (2.17) which is always negative. To find $\text{corr}(\theta_j, \theta_k)$, both of $\text{Var}(\theta_j)$ and $\text{Var}(\theta_k)$ have to be used, where

$$\text{Var}(\theta_j) = \frac{\nu H_j (\nu - \nu H_j)}{\nu^2 (\nu + 1)} = \frac{H_j (1 - H_j)}{\nu + 1}, \quad (2.21)$$

and

$$\text{Var}(\theta_k) = \frac{\nu H_k (\nu - \nu H_k)}{\nu^2 (\nu + 1)} = \frac{H_k (1 - H_k)}{\nu + 1}. \quad (2.22)$$

Then the correlation $\text{corr}(\theta_j, \theta_k)$ is

$$\begin{aligned} \text{corr}(\theta_j, \theta_k) &= \frac{\text{Cov}(\theta_j, \theta_k)}{\sqrt{\text{Var}(\theta_j) \text{Var}(\theta_k)}} = \frac{\frac{H_j (1 - H_k)}{\nu + 1}}{\sqrt{\frac{H_j (1 - H_j)}{\nu + 1} \cdot \frac{H_k (1 - H_k)}{\nu + 1}}} \\ &= \frac{H_j (1 - H_k)}{\sqrt{H_j (1 - H_j) \cdot H_k (1 - H_k)}} = \sqrt{\frac{H_j (1 - H_k)}{H_k (1 - H_j)}}. \end{aligned} \quad (2.23)$$

Bayesian estimates of β_0 and β_1 are computed by finding the expectation of the posterior pdf in (2.12). Unlike the situation in the beta and Dirichlet probability models where expectations can be found by hand, the expected value of (2.12) has to be found computationally. This can be done by estimating $\boldsymbol{\beta}$ using Markov Chain Monte Carlo (MCMC) methods through applying Metropolis-Hastings (MH) sampling algorithm. This algorithm was established by Metropolis *et al* (1953) and being generalized by Hastings (1970). More information about MCMC methods and algorithms can be found in Robert and Casella (1999). In MH algorithm, a proposal pdf, say $q(\cdot)$, is designed to be the stationary distribution that plays the role of the posterior density in (2.12). The proposal function will be used to generate random values for β_0 and β_1 , where every random drawn value will be either accepted or rejected. The means of the accepted values are then computed and considered as estimates of the parameters of interest. Here is a brief description of the MH algorithm:

- Given $\boldsymbol{\beta}^{(t)} = [\beta_0^{(t)}, \beta_1^{(t)}]'$.
- Generate

$$\mathbf{Z} \sim q(\boldsymbol{\beta}^{(t+1)} | \boldsymbol{\beta}^{(t)}). \quad (2.24)$$

- Then let

$$p = \min [\alpha_{MH}, 1], \quad (2.25)$$

where

$$\alpha_{MH} = \frac{\pi(\mathbf{Z}|\mathbf{y}) q(\boldsymbol{\beta}^{(t)}|\mathbf{Z})}{\pi(\boldsymbol{\beta}^{(t)}|\mathbf{y}) q(\mathbf{Z}|\boldsymbol{\beta}^{(t)})}, \quad (2.26)$$

and $\pi(\cdot)$ is the posterior density in (2.12).

- Draw $u \sim U[0, 1]$, so that

$$\boldsymbol{\beta}^{(t+1)} = \begin{cases} \mathbf{Z}, & u \leq p, \\ \boldsymbol{\beta}^{(t)}, & u > p. \end{cases} \quad (2.27)$$

Random values of the regression parameters were drawn independently from $t_{(5)}$ distribution as a proposal pdf. The previous algorithm was applied and iterated over 2000 times after finding the likelihood estimates of β_0 and β_1 . In each iteration, the regression parameters are generated as

$$\beta_j^{(t)} = \hat{\sigma}_{j,LH} \cdot [\text{generated } t_{(5)}\text{-dist value}] + \hat{\beta}_{j,LH}, \quad j = 0, 1, \quad (2.28)$$

where $\hat{\sigma}_{j,LH}$ is the standard error of $\hat{\beta}_j$ (determined by the information matrix in maximum likelihood estimates) and it is fixed over all iterations. Both of $\hat{\sigma}_{j,LH}$ and $\hat{\beta}_{j,LH}$ are required to make a transformation for $t_{(5)}$. Out of 2000 iterations, the first 1000 generations were burned and the other 1000 iterations were considered for estimation purposes. The generation of random draws of regression parameters for the purpose of Bayesian estimation will follow (2.28) in all of the following sections and chapters. Computations are coded using R software package and the code is given in Appendix A.

Table 2.1 presents observed data given in Govindarajulu (1988). The data describes the application of a toxin called retenone to *Macrosiphoniella sanborni*. Estimates (posterior means) of β_0 and β_1 in the logit model are given in tables 2.2 and 2.3 where the term “selected pair” indicates that the prior distribution was defined on the given θ_j and θ_k . Such estimates are close to the maximum likelihood estimates $\hat{\beta}_{0,LH} = -3.223$ and $\hat{\beta}_{1,LH} = 0.606$. Plots of the predicted logistic curves are given in Figure 2.1. Figures 2.2

and 2.3 give the sequences of the generated random values of β_0 and β_1 for the top-left plot of Figure 2.1.

In Figure 2.2, the computational algorithm starts with $\beta_0 = -3.2$, then it keeps accepting or rejecting the next random draws. It can be seen that random draws tend to stabilize by moving around the value of -3.21. A similar thing happens to β_1 in Figure 2.3 where the sequence starts with $\beta_1 = 0.64$ then the adopted random draws tend to stabilize around 0.6.

Table 2.1: Binomial data where the dose is given in mg/l, n is the number of experimented insects and y_{obs} is the number of affected insects.

	Dose	n	y_{obs}
1	0	49	0
2	2.6	50	6
3	3.8	48	16
4	5.1	46	24
5	7.7	49	42
6	10.2	50	44

Table 2.2: β_0 estimates in increasing logit model with $H_1 = 0.15$, $H_2 = 0.3$, $H_3 = 0.45$, $H_4 = 0.55$, $H_5 = 0.7$ and $H_6 = 0.85$.

selected pair	$\nu = 0.5$	$\nu = 1$	$\nu = 2$
θ_1, θ_6	-3.210	-3.221	-3.204
θ_2, θ_5	-3.216	-3.206	-3.229
θ_3, θ_4	-3.227	-3.235	-3.235

Table 2.3: β_1 estimates in increasing logit model with $H_1 = 0.15$, $H_2 = 0.3$, $H_3 = 0.45$, $H_4 = 0.55$, $H_5 = 0.7$ and $H_6 = 0.85$.

selected pair	$\nu = 0.5$	$\nu = 1$	$\nu = 2$
θ_1, θ_6	0.606	0.606	0.603
θ_2, θ_5	0.605	0.602	0.602
θ_3, θ_4	0.604	0.607	0.605

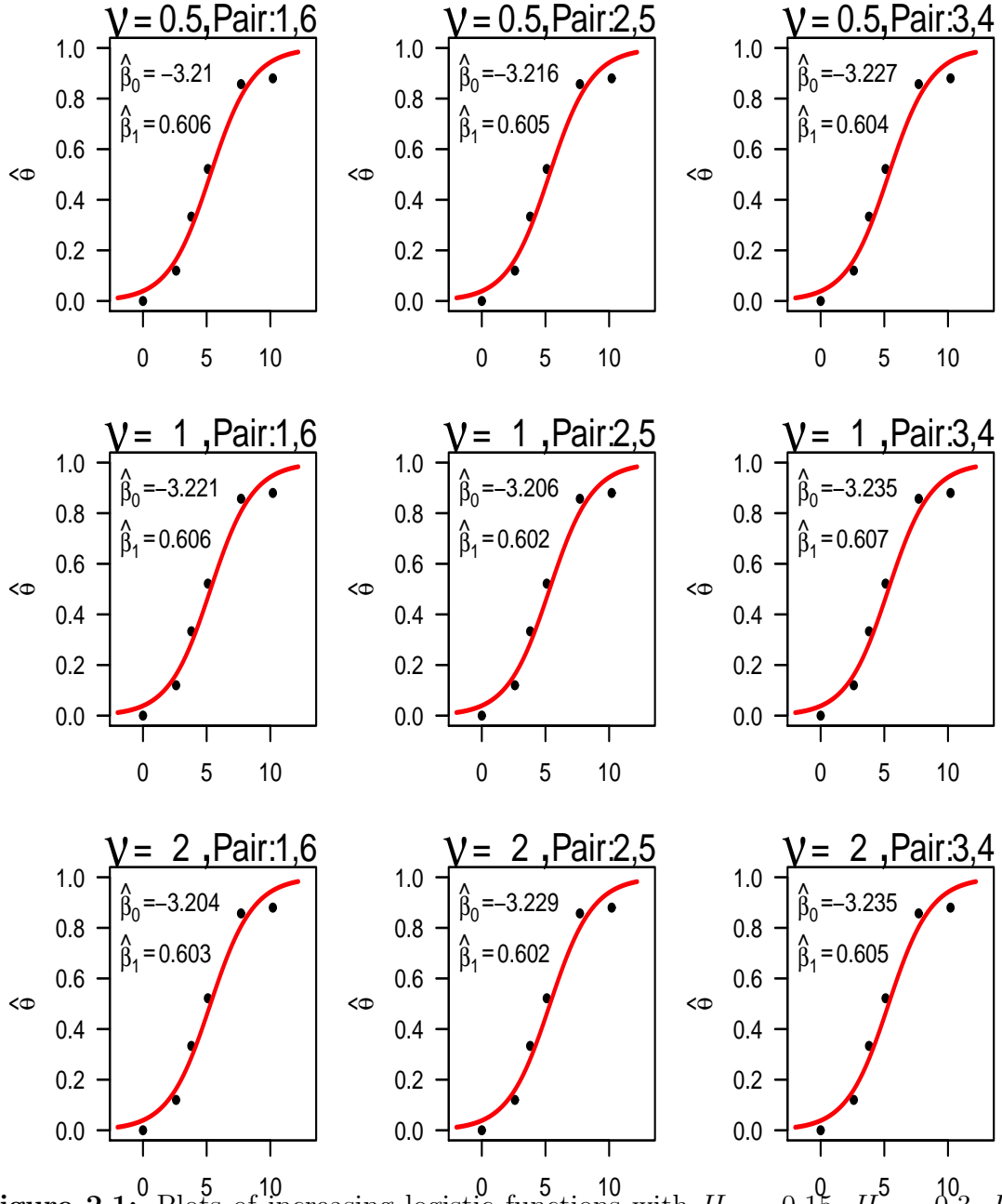


Figure 2.1: Plots of increasing logistic functions with $H_1 = 0.15$, $H_2 = 0.3$, $H_3 = 0.45$, $H_4 = 0.55$, $H_5 = 0.7$ and $H_6 = 0.85$. Top: The selected pairs are θ_1 , θ_6 (top-left), θ_2 , θ_5 (top-middle), θ_3 , θ_4 (top-right) and $\nu = 0.5$. Middle: θ_1 , θ_6 (middle-left), θ_2 , θ_5 (middle-middle), θ_3 , θ_4 (middle-right) and $\nu = 1$. Bottom: θ_1 , θ_6 (bottom-left), θ_2 , θ_5 (bottom-middle), θ_3 , θ_4 (bottom-right) and $\nu = 2$.

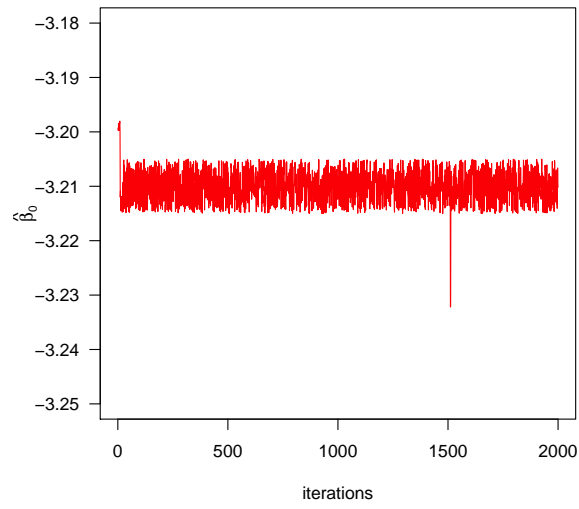


Figure 2.2: Sequence of 2000 random draws of β_0 in MH algorithm of top-left plot in Figure 2.1.

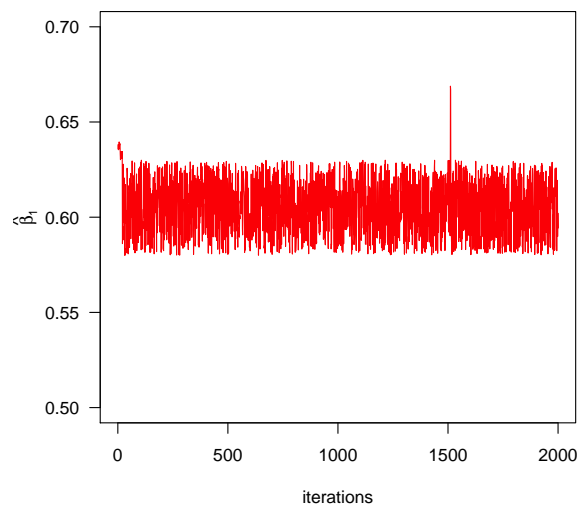


Figure 2.3: Sequence of 2000 random draws of β_1 in MH algorithm of top-left plot in Figure 2.1.

2.2 Decreasing 2-Parameters Logit Model

The decreasing logit model is considered if there is a belief of having lower probability of occurrence with larger value of x_i . The prior distribution assumption is somehow similar in its form to that in (2.2). The prior distribution is

$$(1 - \theta_1, \theta_1 - \theta_2, \dots, \theta_{m-1} - \theta_m, \theta_m) \sim \text{Dir}(\nu - \nu H_1^*, \nu H_1^* - \nu H_2^*, \dots, \nu H_{m-1}^* - \nu H_m^*, \nu H_m^*), \quad (2.29)$$

where after considering the aggregation property and hyperparameters reparametrization, the joint prior distribution for any θ_j, θ_k becomes

$$(1 - \theta_j, \theta_j - \theta_k, \theta_k) \sim \text{Dir}(\nu - \nu H_j^*, \nu H_j^* - \nu H_k^*, \nu H_k^*), \quad (2.30)$$

with joint pdf as

$$\begin{aligned} f(\theta_j, \theta_k) &= \frac{\Gamma(\nu)}{\Gamma(\nu - \nu H_j^*) \Gamma(\nu H_j^* - \nu H_k^*) \Gamma(\nu H_k^*)} \\ &\quad \times (1 - \theta_j)^{\nu - \nu H_j^* - 1} (\theta_j - \theta_k)^{\nu H_j^* - \nu H_k^* - 1} (\theta_k)^{\nu H_k^* - 1}, \\ &\quad 1 > \theta_j > \theta_k > 0. \end{aligned} \quad (2.31)$$

To find $\text{corr}(\theta_j, \theta_k)$,

$$\begin{aligned} \text{Cov}(\theta_j - \theta_k, \theta_k) &= \frac{-(\nu H_j^* - \nu H_k^*) \nu H_k^*}{\nu^2 (\nu + 1)} = \text{Cov}(\theta_j, \theta_k) - \text{Var}(\theta_k) \\ &= \text{Cov}(\theta_j, \theta_k) - \text{Var}(\theta_k). \end{aligned} \quad (2.32)$$

Since

$$(1 - \theta_k) \sim \text{beta}(\nu - \nu H_k^*, \nu H_k^*), \quad (2.33)$$

then

$$\text{Var}(\theta_k) = \frac{\nu H_k^* (\nu - \nu H_k^*)}{\nu^2 (\nu + 1)} = \frac{H_k^* (1 - H_k^*)}{(\nu + 1)}, \quad (2.34)$$

and the same can be found for θ_j

$$\text{Var}(\theta_j) = \frac{\nu H_j^* (\nu - \nu H_j^*)}{\nu^2 (\nu + 1)} = \frac{H_j^* (1 - H_j^*)}{(\nu + 1)}. \quad (2.35)$$

This leads to have

$$\begin{aligned}\text{Cov}(\theta_j, \theta_k) &= \frac{-\nu^2 H_j^* H_k^* + (\nu H_k^*)^2 + \nu^2 H_k^* - (\nu H_k^*)^2}{\nu^2 (\nu + 1)} \\ &= \frac{-\nu^2 H_j^* H_k^* + (\nu H_k^*)^2}{\nu^2 (\nu + 1)} = \frac{H_k^* (1 - H_j^*)}{\nu + 1},\end{aligned}\quad (2.36)$$

which results in

$$\text{corr}(\theta_j, \theta_k) = \frac{\text{Cov}(\theta_j, \theta_k)}{\sqrt{\text{Var}(\theta_j) \text{Var}(\theta_k)}} = \sqrt{\frac{H_k^* (1 - H_j^*)}{H_j^* (1 - H_k^*)}}. \quad (2.37)$$

The joint posterior distribution is induced and written in terms of β_0 and β_1 by following the same method in deriving (2.12) so that

$$\begin{aligned}\pi(\boldsymbol{\beta} | y_1, \dots, y_m) &\propto f(\mathbf{y}, \boldsymbol{\beta}) \\ &= \left\{ \prod_{i=1}^m \left[\binom{n_i}{y_i} \left(\frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right] \right\} \\ &\times \left[\frac{\Gamma(\nu)}{\Gamma(\nu - \nu H_j^*) \Gamma(\nu H_j^* - \nu H_k^*) \Gamma(\nu H_k^*)} \right] \\ &\times \left[\left(1 - \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right)^{\nu - \nu H_j^* - 1} \right. \\ &\quad \cdot \left(\frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} - \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right)^{\nu H_j^* - \nu H_k^* - 1} \left(\frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right)^{\nu H_k^* - 1} \left. \right] \\ &\times |x_k - x_j| \left\{ \left[\frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \right] \left[\frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right] \right\}\end{aligned}\quad (2.38)$$

$$\begin{aligned}
&= \left\{ \prod_{i=1}^m \left[\binom{n_i}{y_i} \left(\frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right] \right\} \\
&\times \left[\frac{\Gamma(\nu)}{\Gamma(\nu - \nu H_j^*) \Gamma(\nu H_j^* - \nu H_k^*) \Gamma(\nu H_k^*)} \right] \\
&\times \left[\left(\frac{1}{1 + e^{\beta_0 + x_j \beta_1}} \right)^{\nu - \nu H_j^*} \left(\frac{e^{\beta_0 + x_j \beta_1} - e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})(1 + e^{\beta_0 + x_k \beta_1})} \right)^{\nu H_k^* - \nu H_j^*} \right. \\
&\quad \left. \cdot \left(\frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right)^{\nu H_k^*} \right] \\
&\times (1 + e^{\beta_0 + x_j \beta_1}) \cdot \left(\frac{(1 + e^{\beta_0 + x_j \beta_1})(1 + e^{\beta_0 + x_k \beta_1})}{e^{\beta_0 + x_j \beta_1} - e^{\beta_0 + x_k \beta_1}} \right) \cdot \left(\frac{1 + e^{\beta_0 + x_k \beta_1}}{e^{\beta_0 + x_k \beta_1}} \right) \\
&\times |x_k - x_j| \left\{ \left[\frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \right] \left[\frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right] \right\} \\
&= \left\{ \prod_{i=1}^m \left[\binom{n_i}{y_i} \left(\frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right] \right\} \\
&\times \left[\frac{\Gamma(\nu)}{\Gamma(\nu - \nu H_j^*) \Gamma(\nu H_j^* - \nu H_k^*) \Gamma(\nu H_k^*)} \right] \\
&\times \left[\left(\frac{1}{1 + e^{\beta_0 + x_j \beta_1}} \right)^{\nu - \nu H_j^*} \left(\frac{e^{\beta_0 + x_j \beta_1} - e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})(1 + e^{\beta_0 + x_k \beta_1})} \right)^{\nu H_k^* - \nu H_j^*} \right. \\
&\quad \left. \cdot \left(\frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right)^{\nu H_k^*} \right] \cdot \left[(x_k - x_j) \left(\frac{e^{\beta_0 + x_j \beta_1}}{e^{\beta_0 + x_j \beta_1} - e^{\beta_0 + x_k \beta_1}} \right) \right]. \quad (2.39)
\end{aligned}$$

Figure 2.4 gives 9 different logistic decreasing fits for data given in Table 2.4. Table 2.4 uses the data given in Table 2.1 but for number of insects that are not affected ($n - y_{obs}$). Tables 2.5 and 2.6 present the numerical estimates of β_0 and β_1 .

Sections 2.1 and 2.2 discussed 2-parameters logit model with a prior of Dirichlet distribution. Section 2.3 presents a new probability model called the logit-normal model. The multivariate logit-normal model is given in Section 2.4. This prepares to introduce the univariate and multivariate forms of imprecise logit-normal model in Chapter 3.

Table 2.4: Binomial data where the dose is given in mg/l, n is the number of experimented insects and $n - y_{obs}$ is the number of insects that are not affected.

	Dose	n	$n - y_{obs}$
1	0	49	49
2	2.6	50	44
3	3.8	48	32
4	5.1	46	22
5	7.7	49	7
6	10.2	50	6

Table 2.5: β_0 estimates in decreasing logit model with $H_1^* = 0.85$, $H_2^* = 0.7$, $H_3^* = 0.55$, $H_4^* = 0.45$, $H_5^* = 0.3$ and $H_6^* = 0.15$.

selected pair	$\nu = 0.5$	$\nu = 1$	$\nu = 2$
θ_1, θ_6	3.216	3.214	3.177
θ_2, θ_5	3.218	3.232	3.174
θ_3, θ_4	3.220	3.214	3.223

Table 2.6: β_1 estimates in decreasing logit model with $H_1^* = 0.85$, $H_2^* = 0.7$, $H_3^* = 0.55$, $H_4^* = 0.45$, $H_5^* = 0.3$ and $H_6^* = 0.15$.

selected pair	$\nu = 0.5$	$\nu = 1$	$\nu = 2$
θ_1, θ_6	-0.599	-0.596	-0.584
θ_2, θ_5	-0.605	-0.606	-0.591
θ_3, θ_4	-0.605	-0.599	-0.607

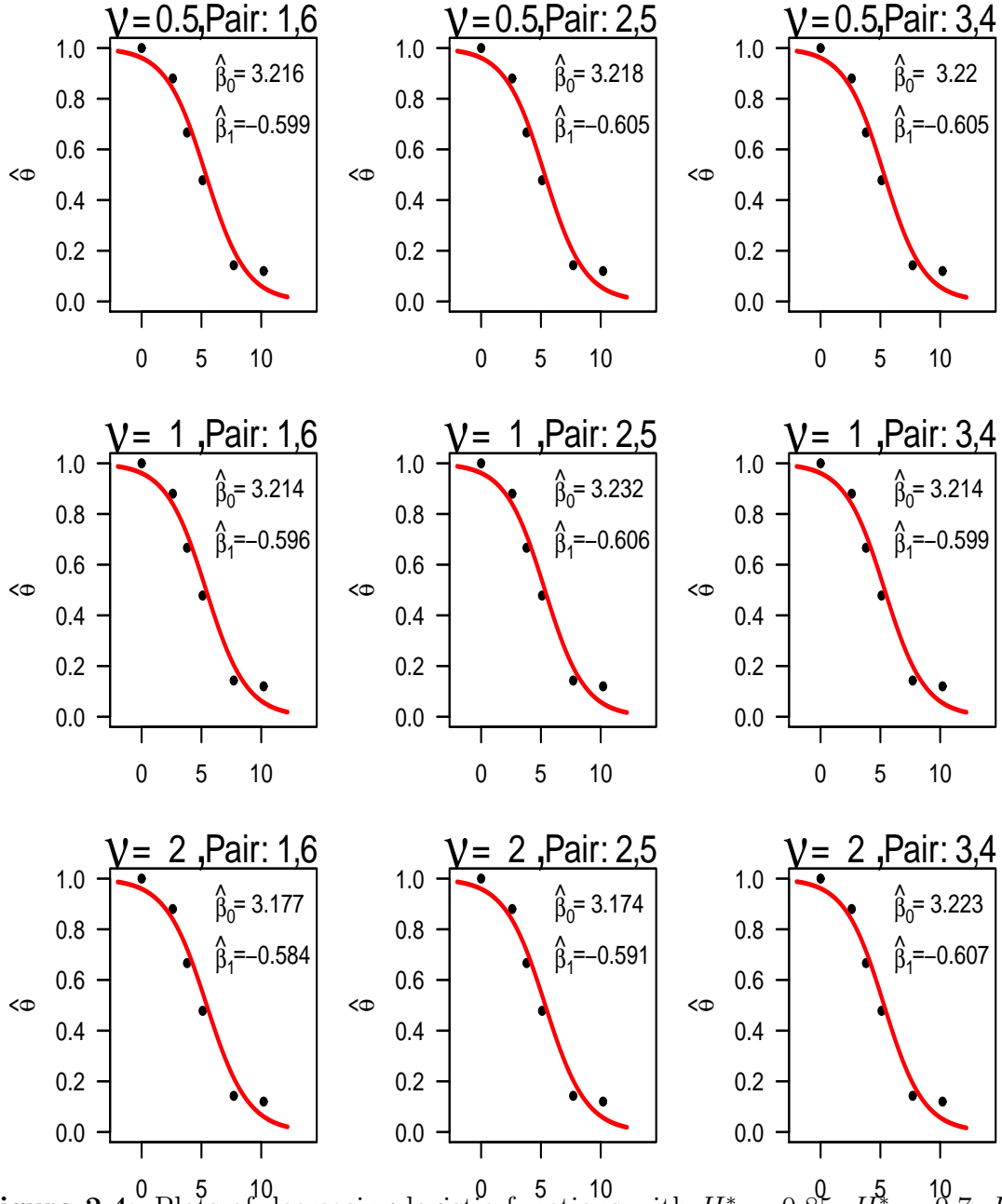


Figure 2.4: Plots of decreasing logistic functions with $H_1^* = 0.85$, $H_2^* = 0.7$, $H_3^* = 0.55$, $H_4^* = 0.45$, $H_5^* = 0.3$ and $H_6^* = 0.15$. Top: The selected pairs are θ_1 , θ_6 (top-left), θ_2 , θ_5 (top-middle), θ_3 , θ_4 (top-right) and $\nu = 0.5$. Middle: θ_1 , θ_6 (middle-left), θ_2 , θ_5 (middle-middle), θ_3 , θ_4 (middle-right) and $\nu = 1$. Bottom: θ_1 , θ_6 (bottom-left), θ_2 , θ_5 (bottom-middle), θ_3 , θ_4 (bottom-right) and $\nu = 2$.

2.3 Logit–Normal Model

The logit-normal pdf is

$$\pi(\theta|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp \left\{ -\frac{\left[\log\left(\frac{\theta}{1-\theta}\right) - \mu \right]^2}{2\sigma^2} \right\} \cdot \left[\frac{1}{\theta(1-\theta)} \right], \quad (2.40)$$

$$0 < \theta < 1, \quad -\infty < \mu < \infty, \quad \sigma^2 > 0,$$

where μ is a location parameter and σ is a scale parameter. The pdf in (2.40) does not have explicit solutions for its mean, mode and variance as mentioned in Mead (1965). The logit-normal model was used as a prior distribution in previous studies and applications. This includes Bloch and Watson (1967) and Leonard (1973). Properties of the logit-normal model can be seen in Aitchison and Shen (1980). Some recent attempts to present methods for computing the moments for the logit-normal distribution are in Frederic and Lad (2008). Figure 2.5 gives plots for logit-normal density.

The relation between the logit-normal distribution and the normal distribution is given in Appendix B. Appendix B shows that the prior and posterior logit-normal distributions (for θ in binomial distribution) belong to an exponential family of distributions.

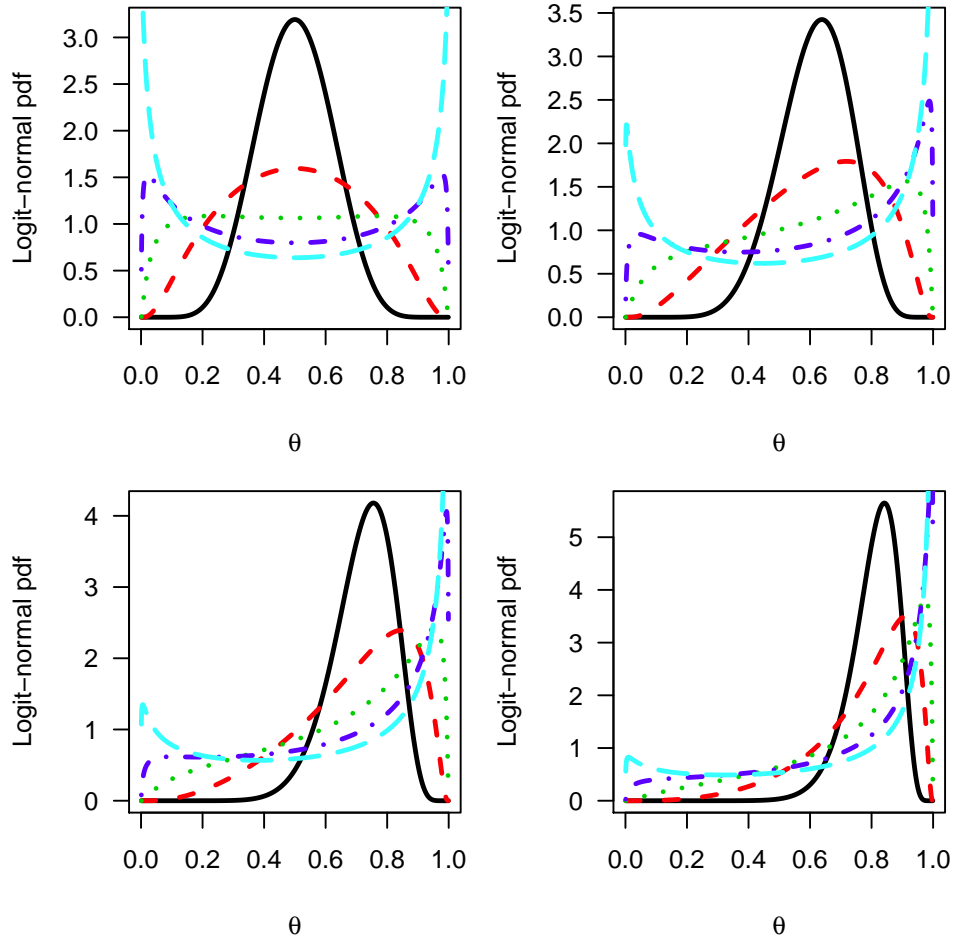


Figure 2.5: Top-left: Plots of logit-normal pdf's with $\mu = 0$, $\sigma = 0.5$ (solid), $\sigma = 1$ (dashed), $\sigma = 1.5$ (dotted), $\sigma = 2$ (dot-dashed), $\sigma = 2.5$ (long-dashed). Top-right: $\mu = 0.5$, $\sigma = 0.5$ (solid), $\sigma = 1$ (dashed), $\sigma = 1.5$ (dotted), $\sigma = 2$ (dot-dashed), $\sigma = 2.5$ (long-dashed). Bottom-left: $\mu = 1$, $\sigma = 0.5$ (solid), $\sigma = 1$ (dashed), $\sigma = 1.5$ (dotted), $\sigma = 2$ (dot-dashed), $\sigma = 2.5$ (long-dashed). Bottom-right: $\mu = 1.5$, $\sigma = 0.5$ (solid), $\sigma = 1$ (dashed), $\sigma = 1.5$ (dotted), $\sigma = 2$ (dot-dashed), $\sigma = 2.5$ (long-dashed).

2.4 Multivariate Logit-Normal Model

The multivariate logit-normal pdf for $\boldsymbol{\theta}_{m \times 1} = [\theta_1, \dots, \theta_m]'$ is

$$\pi(\boldsymbol{\theta} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{m}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \cdot \exp \left\{ -\frac{1}{2} \left[\log \left(\frac{\boldsymbol{\theta}}{1 - \boldsymbol{\theta}} \right) - \boldsymbol{\mu} \right]' \boldsymbol{\Sigma}^{-1} \left[\log \left(\frac{\boldsymbol{\theta}}{1 - \boldsymbol{\theta}} \right) - \boldsymbol{\mu} \right] \right\} \times \prod_{i=1}^m \left[\frac{1}{\theta_i (1 - \theta_i)} \right], \quad (2.41)$$

where

$$\log \left(\frac{\boldsymbol{\theta}}{1 - \boldsymbol{\theta}} \right) = \left[\log \left(\frac{\theta_1}{1 - \theta_1} \right), \dots, \log \left(\frac{\theta_m}{1 - \theta_m} \right) \right]',$$

and

$$\boldsymbol{\mu} = [\mu_1, \dots, \mu_m]',$$

such that $0 < \theta_i < 1$, $-\infty < \mu_i < \infty$, $\boldsymbol{\Sigma}_{m \times m} = [\sigma_{ij}]_{m \times m}$.

Also, $\boldsymbol{\Sigma}_{m \times m} = [\sigma_{ij}]_{m \times m} = \sigma^2 \mathbf{R}_{m \times m}$, where the correlation equation given in (2.23) is considered, for the increasing logistic curve, as

$$\mathbf{R} = \begin{bmatrix} 1 & \sqrt{\frac{H_1(1-H_2)}{H_2(1-H_1)}} & \dots & \sqrt{\frac{H_1(1-H_m)}{H_m(1-H_1)}} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \sqrt{\frac{H_1(1-H_m)}{H_m(1-H_1)}} & \dots & \sqrt{\frac{H_{m-1}(1-H_m)}{H_m(1-H_{m-1})}} & 1 \end{bmatrix}, \quad (2.42)$$

and (2.37) for the decreasing logistic curve as

$$\mathbf{R} = \begin{bmatrix} 1 & \sqrt{\frac{H_2^* (1 - H_1^*)}{H_1^* (1 - H_2^*)}} & \cdot & \dots & \sqrt{\frac{H_m^* (1 - H_1^*)}{H_1^* (1 - H_m^*)}} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sqrt{\frac{H_m^* (1 - H_1^*)}{H_1^* (1 - H_m^*)}} & \dots & \cdot & \sqrt{\frac{H_m^* (1 - H_{m-1}^*)}{H_{m-1}^* (1 - H_m^*)}} & 1 \end{bmatrix}. \quad (2.43)$$

The posterior multivariate logit-normal pdf is

$$\begin{aligned} \pi(\boldsymbol{\theta}|\mathbf{y}) &= \pi(\boldsymbol{\theta}|y_1, \dots, y_m) \\ &= C \cdot \exp \left\{ -\frac{1}{2} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) - (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y}) \right]' \boldsymbol{\Sigma}^{-1} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) - (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y}) \right] \right\} \\ &\quad \times \exp \left\{ \mathbf{1}' \log \left[\frac{(\mathbf{1} - \boldsymbol{\theta})^{\mathbf{n}-1}}{\boldsymbol{\theta}} \right] \right\}, \end{aligned} \quad (2.44)$$

where \mathbf{Y} is a vector of independent binomial random variables.

The prior and posterior multivariate distributions in (2.41) and (2.44) belong to an exponential family of distributions. This is shown in Appendix C. If 2-parameters logit model to be fitted, then a bivariate logit-normal model is assumed as a priori after selecting θ_j and θ_k .

To have a prior bivariate logit-normal model of θ_j and θ_k

$$\begin{aligned} f(\theta_j, \theta_k | \boldsymbol{\mu}_{2 \times 1}, \boldsymbol{\Sigma}_{2 \times 2}) \\ &= \frac{1}{(2\pi)^{|\boldsymbol{\Sigma}_{2 \times 2}|^{\frac{1}{2}}}} \cdot \exp \left\{ -\frac{1}{2} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) - \boldsymbol{\mu} \right]'_{1 \times 2} \boldsymbol{\Sigma}_{2 \times 2}^{-1} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) - \boldsymbol{\mu} \right]_{2 \times 1} \right\} \\ &\times \prod_{i=j,k} \left[\frac{1}{\theta_i (1 - \theta_i)} \right], \end{aligned} \quad (2.45)$$

where

$$\left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) \right]_{2 \times 1} = \begin{bmatrix} \log \left(\frac{\theta_j}{1 - \theta_j} \right) \\ \log \left(\frac{\theta_k}{1 - \theta_k} \right) \end{bmatrix}_{2 \times 1}, \quad [\boldsymbol{\mu}]_{2 \times 1} = \begin{bmatrix} \mu_j \\ \mu_k \end{bmatrix}_{2 \times 1}, \quad (2.46)$$

and

$$\Sigma_{2 \times 2} = \begin{bmatrix} \sigma_j^2 & \sigma_{jk} \\ \sigma_{kj} & \sigma_k^2 \end{bmatrix}_{2 \times 2} = \sigma^2 \mathbf{R}_{2 \times 2} = \sigma^2 \begin{bmatrix} 1 & \rho_{\theta_j, \theta_k} \\ \rho_{\theta_j, \theta_k} & 1 \end{bmatrix}_{2 \times 2}, \quad \sigma_j = \sigma_k, \quad (2.47)$$

where $\rho_{\theta_j, \theta_k}$ can take the form of (2.23) or (2.37), and fixing the value of σ_{jk} results in determining the value of σ in (2.47).

The prior bivariate logit-normal for regression parameters in 2-parameter ImpLogit model is

$$\begin{aligned} \pi(\boldsymbol{\beta}) &= f[\boldsymbol{\theta}(\boldsymbol{\beta})] \left| \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\beta}} \right| = [f(\theta_j, \theta_k)] \left| \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\beta}} \right| = \frac{1}{(2\pi) |\Sigma_{2 \times 2}|^{\frac{1}{2}}}. \\ &\times \exp \left\{ -\frac{1}{2} \left[\log \left(\frac{\boldsymbol{\theta}(\boldsymbol{\beta})}{\mathbf{1} - \boldsymbol{\theta}(\boldsymbol{\beta})} \right) - \boldsymbol{\mu} \right]_{1 \times 2}' \Sigma_{2 \times 2}^{-1} \left[\log \left(\frac{\boldsymbol{\theta}(\boldsymbol{\beta})}{\mathbf{1} - \boldsymbol{\theta}(\boldsymbol{\beta})} \right) - \boldsymbol{\mu} \right]_{2 \times 1} \right\} \\ &\times \prod_{i=1}^2 \left\{ \frac{1}{\theta_i(\beta_0, \beta_1) [1 - \theta_i(\beta_0, \beta_1)]} \right\} \times |x_k - x_j| \left\{ \left[\frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \right] \left[\frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right] \right\}. \end{aligned} \quad (2.48)$$

Then the posterior density of regression parameters is

$$\begin{aligned} \pi(\beta_0, \beta_1 | y_1, \dots, y_m) &\propto f(\mathbf{y}, \boldsymbol{\beta}) = \left\{ \prod_{i=1}^m \left[\binom{n_i}{y_i} \left(\frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right] \right\} \\ &\times \frac{1}{(2\pi) |\Sigma_{2 \times 2}|^{\frac{1}{2}}} \cdot \exp \left\{ -\frac{1}{2} \left[\log \left(\frac{\boldsymbol{\theta}(\boldsymbol{\beta})}{\mathbf{1} - \boldsymbol{\theta}(\boldsymbol{\beta})} \right) - \boldsymbol{\mu} \right]_{1 \times 2}' \Sigma_{2 \times 2}^{-1} \left[\log \left(\frac{\boldsymbol{\theta}(\boldsymbol{\beta})}{\mathbf{1} - \boldsymbol{\theta}(\boldsymbol{\beta})} \right) - \boldsymbol{\mu} \right]_{2 \times 1} \right\} \\ &\times \prod_{i=j,k} \left\{ \frac{1}{\theta_i(\beta_0, \beta_1) [1 - \theta_i(\beta_0, \beta_1)]} \right\} \\ &\times |x_k - x_j| \left\{ \left[\frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \right] \left[\frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right] \right\}, \end{aligned}$$

$$\begin{aligned}
&= \left\{ \prod_{i=1}^m \left[\binom{n_i}{y_i} \left(\frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right] \right\} \\
&\times \frac{1}{(2\pi)^{|\Sigma_{2 \times 2}|^{\frac{1}{2}}}} \cdot \exp \left\{ -\frac{1}{2} \left[\log \left(\frac{\boldsymbol{\theta}(\boldsymbol{\beta})}{\mathbf{1} - \boldsymbol{\theta}(\boldsymbol{\beta})} \right) - \boldsymbol{\mu} \right]_{1 \times 2}' \Sigma_{2 \times 2}^{-1} \left[\log \left(\frac{\boldsymbol{\theta}(\boldsymbol{\beta})}{\mathbf{1} - \boldsymbol{\theta}(\boldsymbol{\beta})} \right) - \boldsymbol{\mu} \right]_{2 \times 1} \right\} \\
&\times \left[\frac{(1 + e^{\beta_0 + x_j \beta_1})^2}{e^{\beta_0 + x_j \beta_1}} \right] \cdot \left[\frac{(1 + e^{\beta_0 + x_k \beta_1})^2}{e^{\beta_0 + x_k \beta_1}} \right] \\
&\times |x_k - x_j| \left\{ \left[\frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \right] \left[\frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right] \right\} \\
&= \left\{ \prod_{i=1}^m \left[\binom{n_i}{y_i} \left(\frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right] \right\} \\
&\times \frac{1}{(2\pi)^{|\Sigma_{2 \times 2}|^{\frac{1}{2}}}} \cdot \exp \left\{ -\frac{1}{2} \left[\log \left(\frac{\boldsymbol{\theta}(\boldsymbol{\beta})}{\mathbf{1} - \boldsymbol{\theta}(\boldsymbol{\beta})} \right) - \boldsymbol{\mu} \right]_{1 \times 2}' \Sigma_{2 \times 2}^{-1} \left[\log \left(\frac{\boldsymbol{\theta}(\boldsymbol{\beta})}{\mathbf{1} - \boldsymbol{\theta}(\boldsymbol{\beta})} \right) - \boldsymbol{\mu} \right]_{2 \times 1} \right\} \\
&\times |x_k - x_j|, \tag{2.49}
\end{aligned}$$

where

$$\left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) \right]_{2 \times 1} = \begin{bmatrix} \beta_0 + x_j \beta_1 \\ \beta_0 + x_k \beta_1 \end{bmatrix}_{2 \times 1}, \quad [\boldsymbol{\mu}]_{2 \times 1} = \begin{bmatrix} \mu_j \\ \mu_k \end{bmatrix}_{2 \times 1}. \tag{2.50}$$

Figure 2.6 shows plots of increasing logistic regression model under the bivariate logit-normal prior distribution for data in Table 2.1. Computed estimates for β_0 and β_1 are given in tables 2.7 and 2.8.

Table 2.7: β_0 estimates in increasing logit model with $H_1 = 0.15$, $H_2 = 0.3$, $H_3 = 0.45$, $H_4 = 0.55$, $H_5 = 0.7$ and $H_6 = 0.85$, $\mu_i = 0$ ($\mu_j = \mu_k$), and $\sigma_i = 0.5, 1$ and 1.5 ($\sigma_j = \sigma_k$), $i = j, k$.

selected pair	$\sigma_i = 0.5$	$\sigma_i = 1$	$\sigma_i = 1.5$
θ_1, θ_6	-3.233	-3.227	-3.234
θ_2, θ_5	-3.241	-3.236	-3.209
θ_3, θ_4	-3.236	-3.232	-3.213

Table 2.8: β_1 estimates in increasing logit model with $H_1 = 0.15$, $H_2 = 0.3$, $H_3 = 0.45$, $H_4 = 0.55$, $H_5 = 0.7$ and $H_6 = 0.85$, $\mu_i = 0$ ($\mu_j = \mu_k$), and $\sigma_i = 0.5, 1$ and 1.5 ($\sigma_j = \sigma_k$), $i = j, k$.

selected pair	$\sigma_i = 0.5$	$\sigma_i = 1$	$\sigma_i = 1.5$
θ_1, θ_6	0.602	0.607	0.605
θ_2, θ_5	0.613	0.605	0.602
θ_3, θ_4	0.607	0.603	0.609

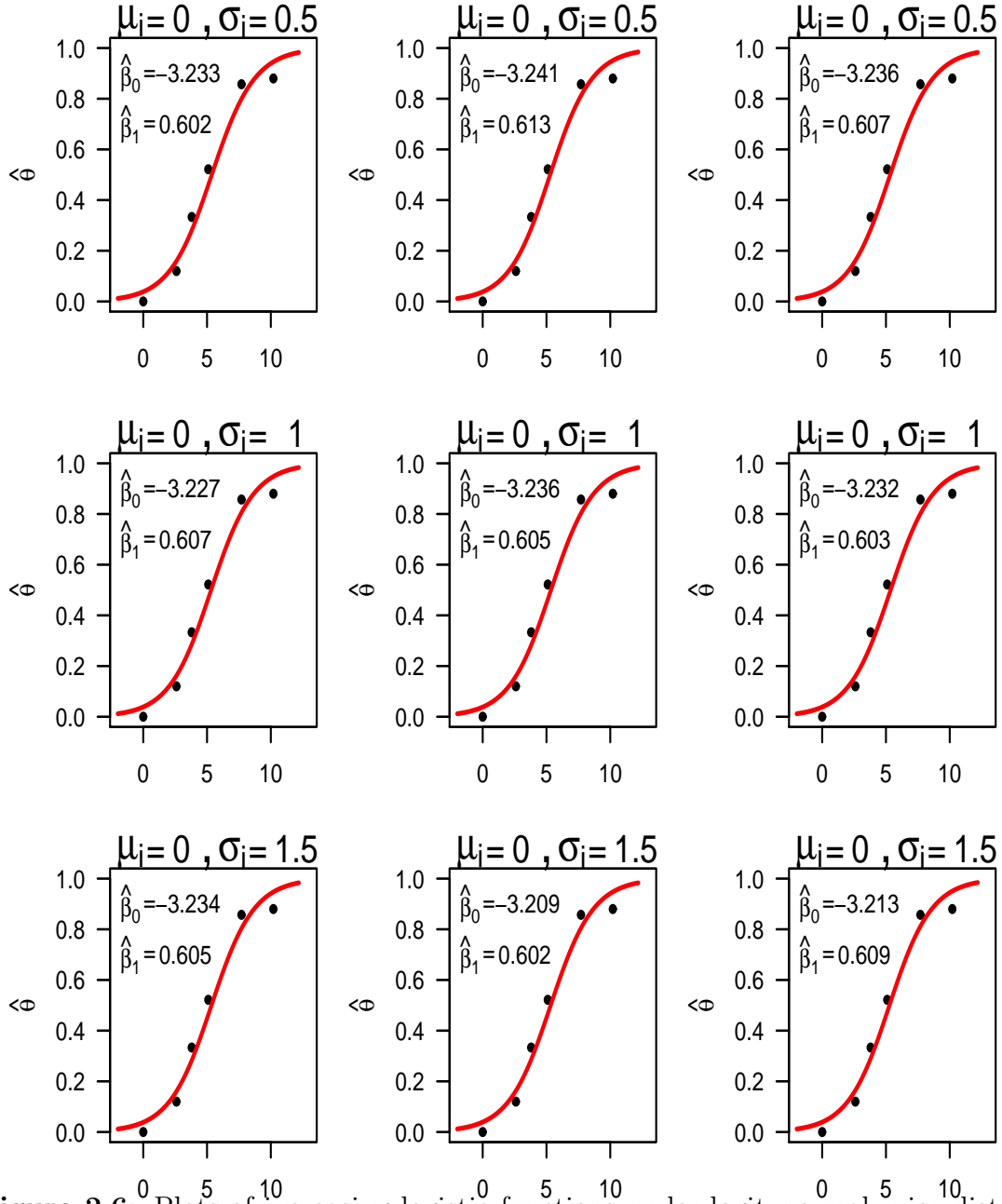


Figure 2.6: Plots of increasing logistic functions under logit-normal prior distribution with $H_1 = 0.15$, $H_2 = 0.3$, $H_3 = 0.45$, $H_4 = 0.55$, $H_5 = 0.7$ and $H_6 = 0.85$. Top: A plot of increasing logistic functions with selected pair θ_1, θ_6 (top-left), θ_2, θ_5 (top-middle), θ_3, θ_4 (top-right) and $\sigma_i = 0.5$. Middle: θ_1, θ_6 (middle-left), θ_2, θ_5 (middle-middle), θ_3, θ_4 (middle-right) and $\sigma_i = 1$. Bottom: θ_1, θ_6 (bottom-left), θ_2, θ_5 (bottom-middle), θ_3, θ_4 (bottom-right) and $\sigma_i = 2$.

CHAPTER 3

IMPLGIT MODEL

This chapter extends Chapter 2 to involve probabilistic imprecision in logit model regression parameters. Using imprecise priors in logit model creates a novel imprecise logistic regression model. The imprecise logistic regression model is suggested to be called as the “ImpLogit model”. ImpLogit model performs a set of Bayesian logit fits with interval estimates for regression parameters.

This chapter establishes a general framework for ImpLogit model. The prior distribution in the logit model will be expressed imprecisely by re-parameterizing the hyperparameters of interest in a way similar to that made for the beta and Dirichlet models in Section 1.7. ImpLogit is the first step in establishing a certain criterion for the analysis of imprecise logistic regression. ImpLogit model will focus on computing interval estimates of the regression parameters.

The 2-parameter ImpLogit model is presented simply in Section 3.1. Then Section 3.2 explores the relation between any selected pair of binomial parameters, θ_j, θ_k in ImpLogit model and the regression parameters, β_0 and β_1 . Details of the structure of increasing and decreasing cases of ImpLogit model are given in sections 3.3 and 3.4. Section 3.5 builds the ImpLogit model with a mixture of beliefs of increasing and decreasing logistic functions where there are no assumptions on the values of β_1 which gives a general manipulation of the ImpLogit model. Section 3.6 considers fitting the ImpLogit model using the imprecise bivariate logit normal model as an imprecise prior. Finally, Section 3.7 fits ImpLogit under IDM with a small sample.

3.1 Imprecision in 2-Parameter Logit Model

ImpLogit model requires a set of logistic functions to be fitted so that a set of estimates for each probability of occurrence, $\hat{\theta}_i$ for all i , are found. The logistic functions in ImpLogit model that correspond to the interval endpoints of β_0 and β_1 are given as

$$\theta_i = \theta(x_i) = \frac{e^{\underline{\beta}_0 + x_i \underline{\beta}_1}}{1 + e^{\underline{\beta}_0 + x_i \underline{\beta}_1}} \implies \log \left(\frac{\theta_i}{1 - \theta_i} \right) = \underline{\beta}_0 + x_i \underline{\beta}_1, \quad (3.1)$$

$$\text{and } \theta_i = \theta(x_i) = \frac{e^{\bar{\beta}_0 + x_i \bar{\beta}_1}}{1 + e^{\bar{\beta}_0 + x_i \bar{\beta}_1}} \implies \log \left(\frac{\theta_i}{1 - \theta_i} \right) = \bar{\beta}_0 + x_i \bar{\beta}_1, \quad (3.2)$$

$$\text{where } \beta_j \in [\underline{\beta}_j, \bar{\beta}_j], \quad j = 0, 1, \text{ and } i = 1, \dots, m.$$

Figure 3.1 describes ImpLogit model shape where logistic curves that correspond to endpoints of β_0 and β_1 in (3.1) and (3.2) are plotted. The two plotted logistic curves intersect at some fixed covariate point. This means that lower and upper values of regression parameters do not correspond to lower and upper logistic curves. That is, having

$$\theta_i = \frac{e^{\underline{\beta}_0 + x_i \underline{\beta}_1}}{1 + e^{\underline{\beta}_0 + x_i \underline{\beta}_1}} \quad \text{and} \quad \theta_i^* = \frac{e^{\bar{\beta}_0 + x_i \bar{\beta}_1}}{1 + e^{\bar{\beta}_0 + x_i \bar{\beta}_1}},$$

does not necessarily imply that $\theta_i < \theta_i^*, \forall i$. This can be figured out in another way where Figure 3.1 shows that regression parameters lower and upper points do not necessarily match with lower and upper points of θ_i at a fixed $x_i, \forall i$. That is, $\underline{\beta}_0$ and $\underline{\beta}_1$ may correspond to points that belong to the interval value of θ_i , but not the lower value $\underline{\theta}_i$ at x_i . This is an interesting point that deserves to be investigated. Therefore, the relation between β_0, β_1 and θ_j, θ_k can provide an understanding of the behaviour of the ImpLogit model. Such relation will be studied in Section 3.2.

Figure 3.2 gives sets of logistic function curves. The curves are plotted under different combinations of values β_0 and β_1 selected from determined interval values. Generally, the intervals of θ_i become shorter at extreme fixed values of x_i , but longer around $x_i = 0$. In Figure 3.2, the top left and right plots seem more spread and imprecise compared to the bottom ones. This is due to the length of the interval value of β_1 which is larger in both top plots. This does not mean that β_0 interval value has a negligible effect on the

ImpLogit fit imprecision, but β_1 seems to play a stronger role in determining the spread of the whole set of logistic function plots. This comes from comparing both of the top and bottom plots in Figure 3.2.

Figure 3.2 can provide information about the role of β_0 on ImpLogit imprecision. In both top plots, where β_1 interval endpoints and length are similar, β_0 made a significant change of imprecision in the whole fit. In the top-left plot there is $\beta_0 \in [0.5, 0.7]$ with interval length of $0.7 - 0.5 = 0.2$, and in the top-right plot there is $\beta_0 \in [-0.5, 0.7]$ with interval length $0.7 - (-0.5) = 1.2$. It can be seen that the whole spread in the top-right plot is more than that in the top-left one. However, this situation is not the same in the bottom plots. Despite that β_0 interval endpoints and lengths are similar to the top plots, but β_1 has a similar length with larger interval endpoints. Bottom plots show that larger β_1 endpoints may result in less effect than β_0 parameter on the imprecision in ImpLogit model fit.

Figure 3.2 shows that both logistic curves with lower and upper regression parameters values tend to coincide for extreme values of the covariate x . This helps to know where can imprecision be reduced in ImpLogit model. This can be useful in putting a general design to make logistic curves as close as possible.

In Figure 3.2, the green color does not pass the blue and red curves over some intervals of the covariate x . It is interesting to find where can this happen, for an increasing or decreasing 2-parameter logistic curves, by solving

$$\frac{e^{\underline{\beta}_0 + x\underline{\beta}_1}}{1 + e^{\underline{\beta}_0 + x\underline{\beta}_1}} < \theta < \frac{e^{\bar{\beta}_0 + x\bar{\beta}_1}}{1 + e^{\bar{\beta}_0 + x\bar{\beta}_1}}, \quad (3.3)$$

and

$$\frac{e^{\underline{\beta}_0 + x\underline{\beta}_1}}{1 + e^{\underline{\beta}_0 + x\underline{\beta}_1}} > \theta > \frac{e^{\bar{\beta}_0 + x\bar{\beta}_1}}{1 + e^{\bar{\beta}_0 + x\bar{\beta}_1}}. \quad (3.4)$$

Inequality (3.3) is re-written as

$$\frac{1}{1 + e^{-(\underline{\beta}_0 + x\underline{\beta}_1)}} < \frac{1}{1 + e^{-(\beta_0 + x\beta_1)}} < \frac{1}{1 + e^{-(\bar{\beta}_0 + x\bar{\beta}_1)}},$$

\implies

$$1 + e^{-(\beta_0 + x\beta_1)} < 1 + e^{-(\underline{\beta}_0 + x\underline{\beta}_1)}, \quad (3.5)$$

and

$$1 + e^{-(\bar{\beta}_0 + x\bar{\beta}_1)} < 1 + e^{-(\beta_0 + x\beta_1)}. \quad (3.6)$$

Then solving (3.5) gives

$$-(\beta_0 + x\beta_1) < -(\underline{\beta}_0 + x\underline{\beta}_1),$$

\implies

$$\underline{\beta}_0 - \beta_0 < x(\beta_1 - \underline{\beta}_1) \implies \frac{\underline{\beta}_0 - \beta_0}{\beta_1 - \underline{\beta}_1} < x, \quad (3.7)$$

and solving (3.6) gives

$$-(\bar{\beta}_0 + x\bar{\beta}_1) < -(\beta_0 + x\beta_1),$$

\implies

$$\beta_0 + x\beta_1 < \bar{\beta}_0 + x\bar{\beta}_1 \implies x(\beta_1 - \bar{\beta}_1) < \bar{\beta}_0 - \beta_0,$$

\implies

$$x > \frac{\bar{\beta}_0 - \beta_0}{\beta_1 - \bar{\beta}_1}. \quad (3.8)$$

Finally, merging (3.7) and (3.8) gives

$$\boxed{\min\left(\frac{\underline{\beta}_0 - \beta_0}{\beta_1 - \underline{\beta}_1}, \frac{\bar{\beta}_0 - \beta_0}{\beta_1 - \bar{\beta}_1}\right) < x.} \quad (3.9)$$

Now, inequality (3.4) is solved in a similar way to (3.3). Then (3.5) and (3.6) become

$$1 + e^{-(\beta_0 + x\beta_1)} > 1 + e^{-(\underline{\beta}_0 + x\underline{\beta}_1)}, \quad (3.10)$$

and

$$1 + e^{-(\bar{\beta}_0 + x\bar{\beta}_1)} > 1 + e^{-(\beta_0 + x\beta_1)}. \quad (3.11)$$

Both of (3.10) and (3.11) give

$$x < \frac{\beta_0 - \underline{\beta}_0}{\underline{\beta}_1 - \beta_1}, \quad (3.12)$$

and

$$x < \frac{\beta_0 - \bar{\beta}_0}{\bar{\beta}_1 - \beta_1}, \quad (3.13)$$

then

$$x < \max \left(\frac{\beta_0 - \underline{\beta}_0}{\underline{\beta}_1 - \beta_1}, \frac{\beta_0 - \bar{\beta}_0}{\bar{\beta}_1 - \beta_1} \right). \quad (3.14)$$

The point of intersection of logistic curves under lower and upper values of regression parameters is

$$\begin{aligned} \frac{1}{1 + e^{-(\beta_0 + x\underline{\beta}_1)}} &= \frac{1}{1 + e^{-(\bar{\beta}_0 + x\bar{\beta}_1)}} \\ \Rightarrow e^{-(\beta_0 + x\underline{\beta}_1)} &= e^{-(\bar{\beta}_0 + x\bar{\beta}_1)} \\ \beta_0 + x\underline{\beta}_1 &= \bar{\beta}_0 + x\bar{\beta}_1 \\ \underline{\beta}_0 - \bar{\beta}_0 &= x(\bar{\beta}_1 - \underline{\beta}_1), \end{aligned}$$

which becomes

$$x = \frac{\underline{\beta}_0 - \bar{\beta}_0}{\bar{\beta}_1 - \underline{\beta}_1}, \quad (3.15)$$

where x in (3.15) is always having a negative value. This can be figured out in figures 3.1 and 3.2. The same is found for inequality (3.14) where x is always negative valued.

If $x \geq 0$ then it can be seen that

$$\frac{e^{\beta_0 + x\underline{\beta}_1}}{1 + e^{\beta_0 + x\underline{\beta}_1}} < \frac{e^{\bar{\beta}_0 + x\bar{\beta}_1}}{1 + e^{\bar{\beta}_0 + x\bar{\beta}_1}}, \quad \forall x \geq 0. \quad (3.16)$$

This means that with $\beta_0 \in (\underline{\beta}_0, \bar{\beta}_0)$ and $\beta_1 \in (\underline{\beta}_1, \bar{\beta}_1)$, then $\underline{\theta} < \theta < \bar{\theta}$ as in (3.3) if x belongs to (3.9), and $\underline{\theta} > \theta > \bar{\theta}$ as in (3.4) if x belongs to (3.14).

Figures 3.3, 3.4 and 3.5 give a better idea to visualize the interesting relation. Figure 3.3 shows clearly that all combinations of β_0 and β_1 , which form a square in the top-left plot, do not necessarily transmit to a square in the rest of plots. A similar thing can be seen in figures 3.4 and 3.5.

Note that the covariate x_i in (3.9), (3.14) and (3.16) can have either negative or nonnegative values. Now, define

$$h = \min_{i=1, \dots, n} x_i, \quad (3.17)$$

and set

$$\tilde{x}_i = x_i + |h|, \quad (3.18)$$

so that all \tilde{x}_i values are nonnegative. This will affect the parameters in logit model as

$$\tilde{\eta}_i = \tilde{\beta}_0 + \tilde{x}_i \tilde{\beta}_1, \quad (3.19)$$

such that

$$\eta_i = \beta_0 + x_i \beta_1, \quad (3.20)$$

which means that (3.19) and (3.20) become equal if

$$\tilde{\beta}_1 = \beta_1, \quad (3.21)$$

and

$$\tilde{\beta}_0 = \beta_0 - \beta_1 |h|. \quad (3.22)$$

Having the regression parameters as in (3.21) and (3.22) comes by having nonnegative values of the covariate \tilde{x}_i which falls under the case of (3.16). This means that lower and upper logistic curves will correspond to lower and upper values of the transformed regression parameters in (3.21) and (3.22) over the \tilde{x}_i values in (3.18).

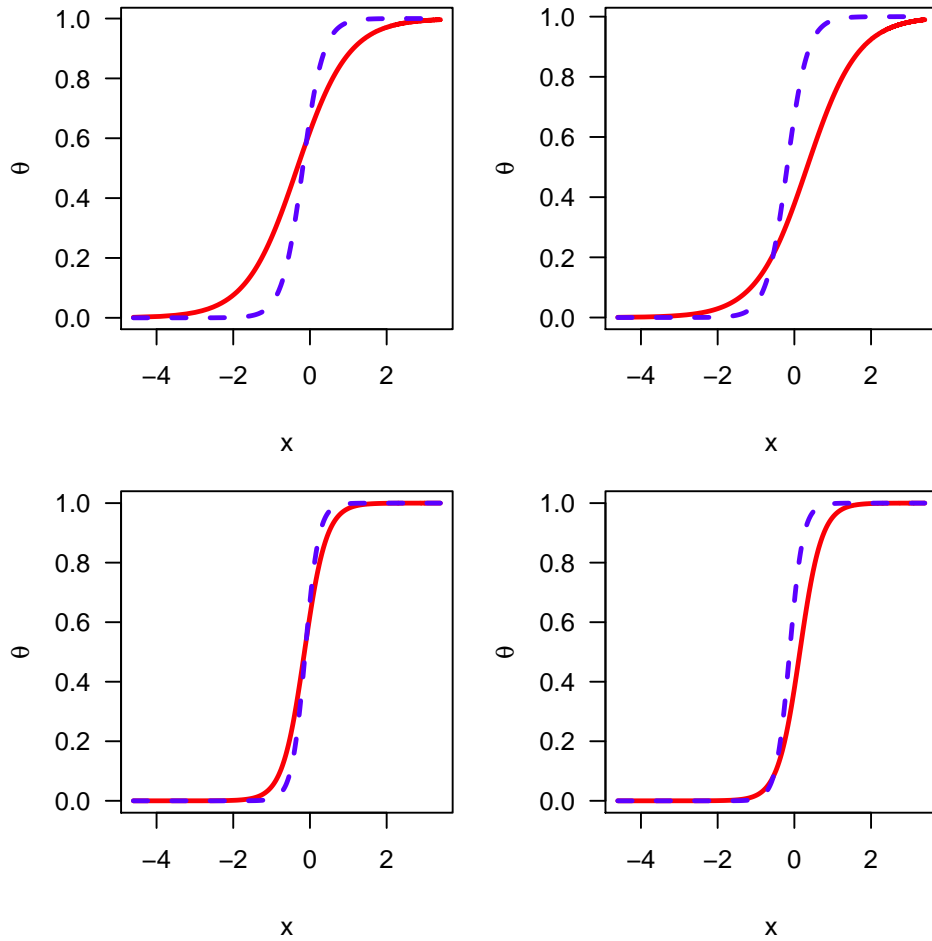


Figure 3.1: Top-Left: Plots of logistic functions with $\beta_0 = 0.5$, $\beta_1 = 1.5$ (solid), $\beta_0 = 0.7$, $\beta_1 = 3.7$ (dashed). Top-Right: $\beta_0 = -0.5$, $\beta_1 = 1.5$ (solid), $\beta_0 = 0.7$, $\beta_1 = 3.7$ (dashed). Bottom-Left: $\beta_0 = 0.5$, $\beta_1 = 3.5$ (solid), $\beta_0 = 0.7$, $\beta_1 = 5.7$ (dashed). Bottom-Right: $\beta_0 = -0.5$, $\beta_1 = 3.5$ (solid), $\beta_0 = 0.7$, $\beta_1 = 5.7$ (dashed).

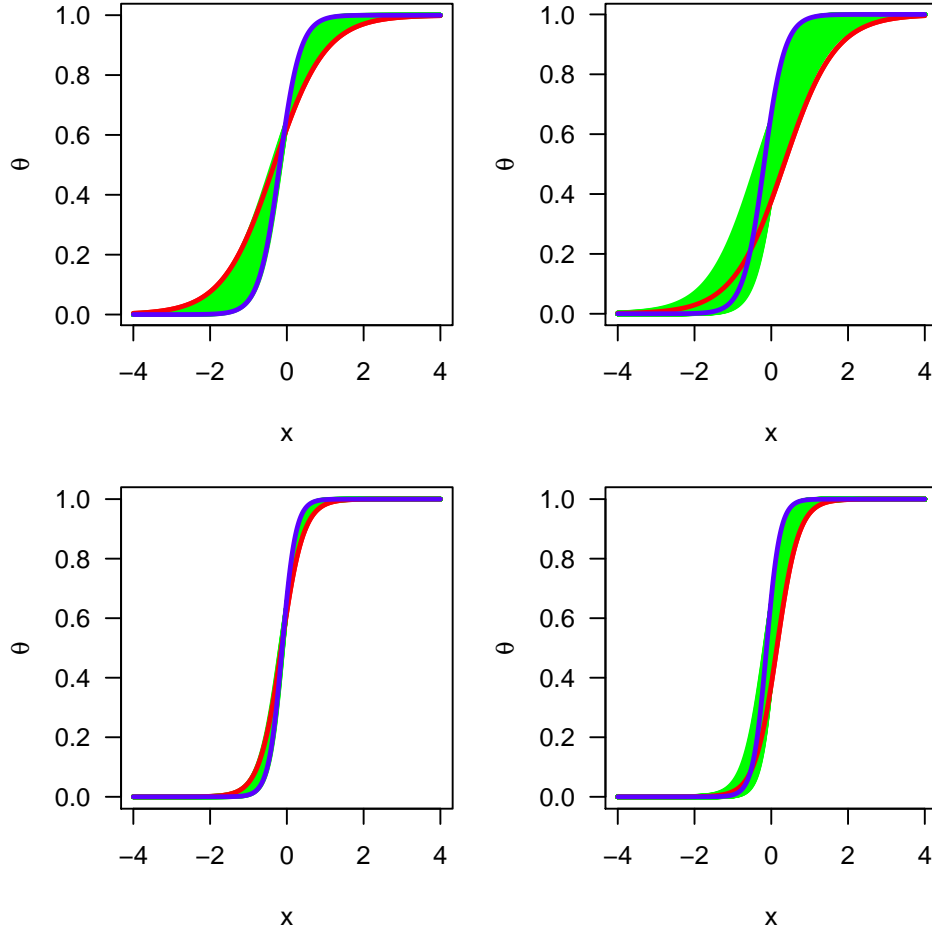


Figure 3.2: The red curve refers to the lower logistic curve with lower values of regression parameters. The blue curve refers to the upper logistic curve with upper values of regression parameters, while the green area is for logistic curves with in between values of regression parameters. Top-Left: Sets of logistic functions with $\beta_0 \in [0.5, 0.7]$ and $\beta_1 \in [1.5, 3.7]$. Top-Right: $\beta_0 \in [-0.5, 0.7]$ and $\beta_1 \in [1.5, 3.7]$. Bottom-Left: $\beta_0 \in [0.5, 0.7]$ and $\beta_1 \in [3.5, 5.7]$. Bottom-Right: $\beta_0 \in [-0.5, 0.7]$ and $\beta_1 \in [3.5, 5.7]$.

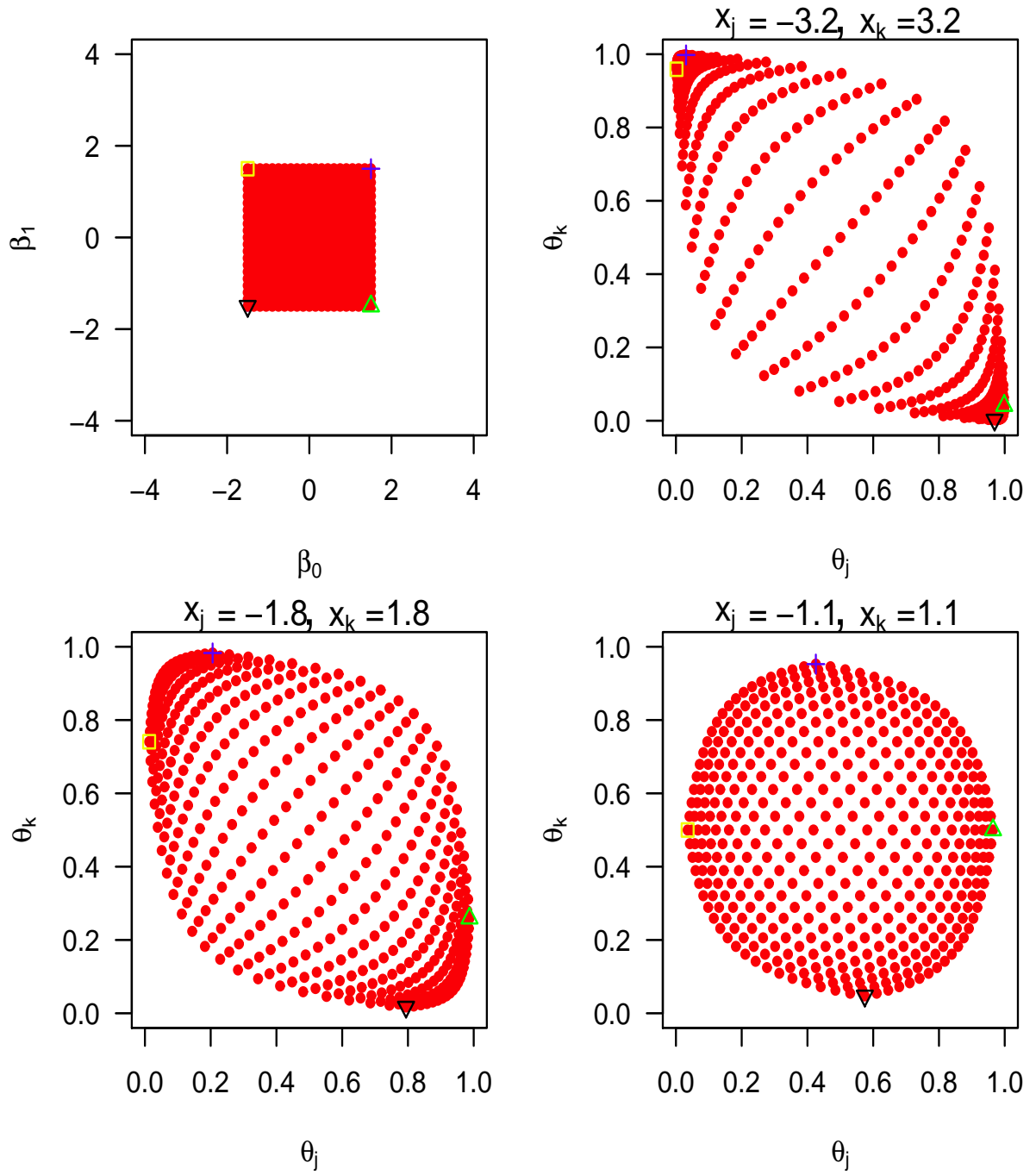


Figure 3.3: Transformation plots from β_0, β_1 to θ_j, θ_k under symmetric x_j and x_k . Top-left: A plot of β_0 vs β_1 . Top-right: A plot of θ_j vs θ_k that corresponds to combinations of β_0 and β_1 with $x_j = -3.2$ and $x_k = 3.2$. Bottom-left: $x_j = -1.8$ and $x_k = 1.8$. Bottom-right: $x_j = -1.1$ and $x_k = 1.1$.

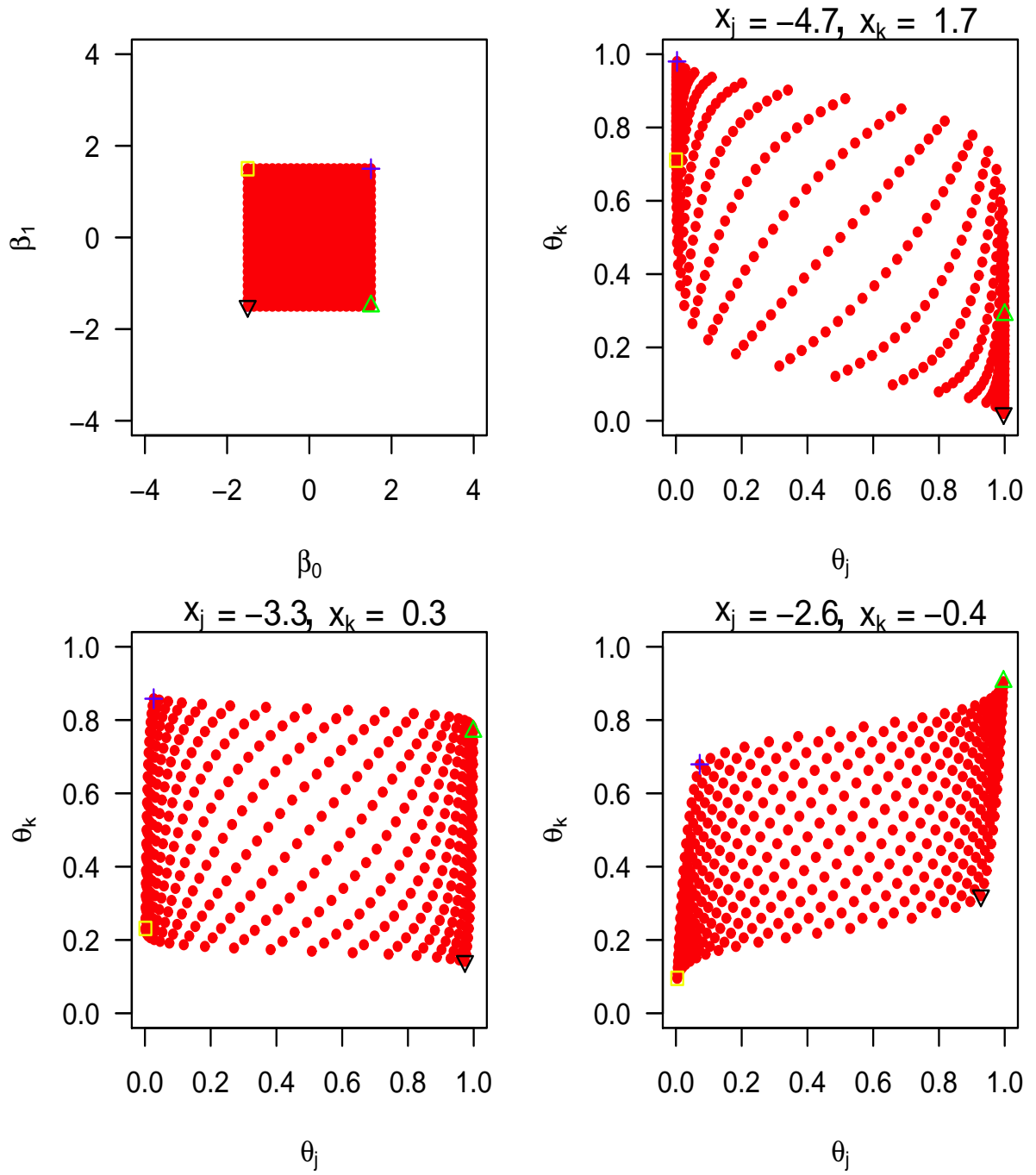


Figure 3.4: Transformation plots from β_0, β_1 to θ_j, θ_k under left-shifted x_j and x_k . Top-left: A plot of β_0 vs β_1 . Top-right: A plot of θ_j vs θ_k that corresponds to combinations of β_0 and β_1 with $x_j = -4.7$ and $x_k = 1.7$. Bottom-left: $x_j = -3.3$ and $x_k = 0.3$. Bottom-right: $x_j = -2.6$ and $x_k = -0.4$.

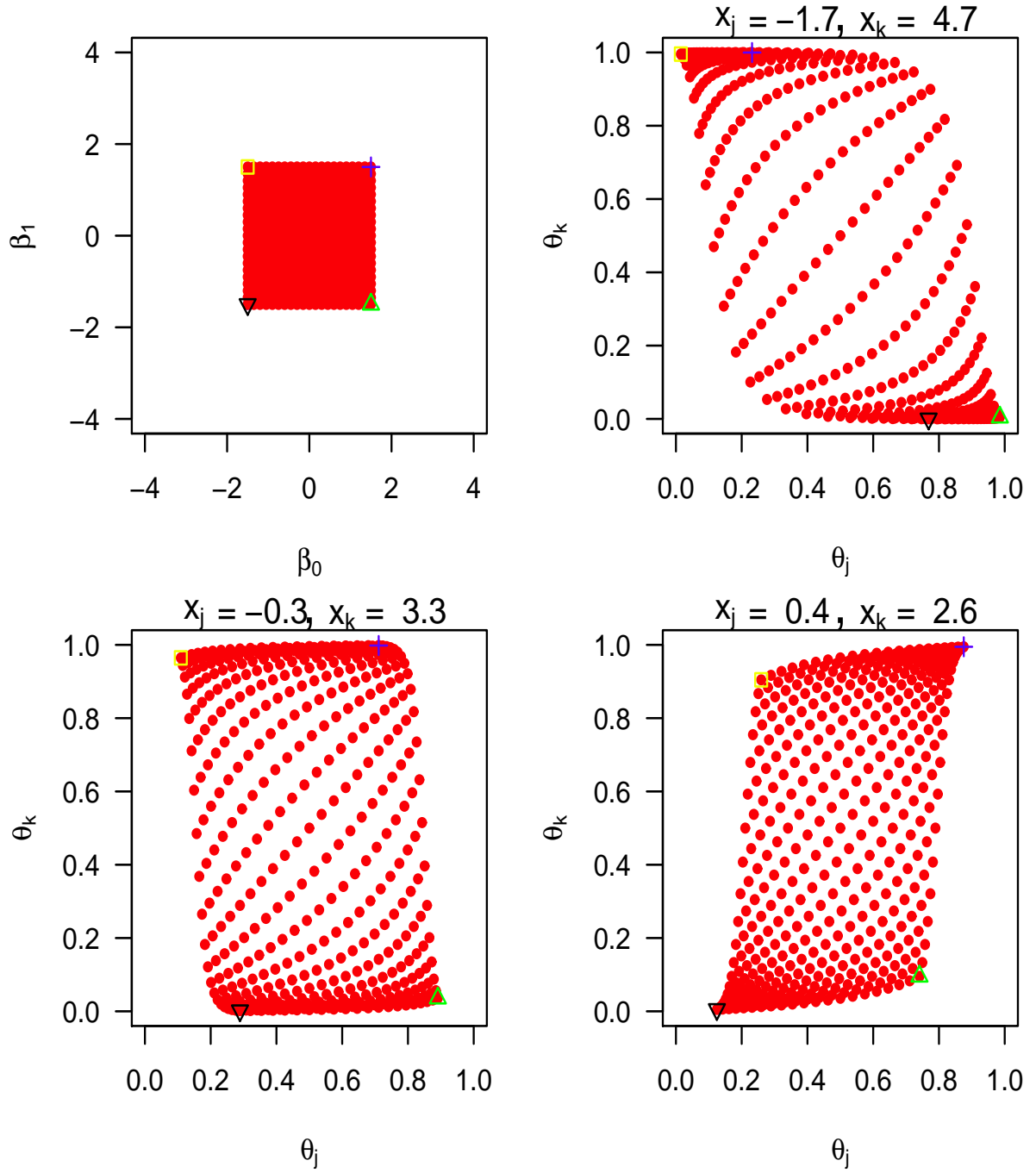


Figure 3.5: Transformation plots from β_0, β_1 to θ_j, θ_k under right-shifted x_j and x_k . Top-left: A plot of β_0 vs β_1 . Top-right: A plot of θ_j vs θ_k that corresponds to combinations of β_0 and β_1 with $x_j = -1.7$ and $x_k = 4.7$. Bottom-left: $x_j = -0.3$ and $x_k = 3.3$. Bottom-right: $x_j = 0.4$ and $x_k = 2.6$.

3.2 Transmission from θ to β in 2-Parameter ImpLogit Model

The mathematical relation that maps sets of $\beta = [\beta_0, \beta_1]'$ to sets of $\theta = [\theta_j, \theta_k]'$ and vice versa is important to know. Figure 3.6 divides the plotted area of θ_j vs θ_k into 4 regions.

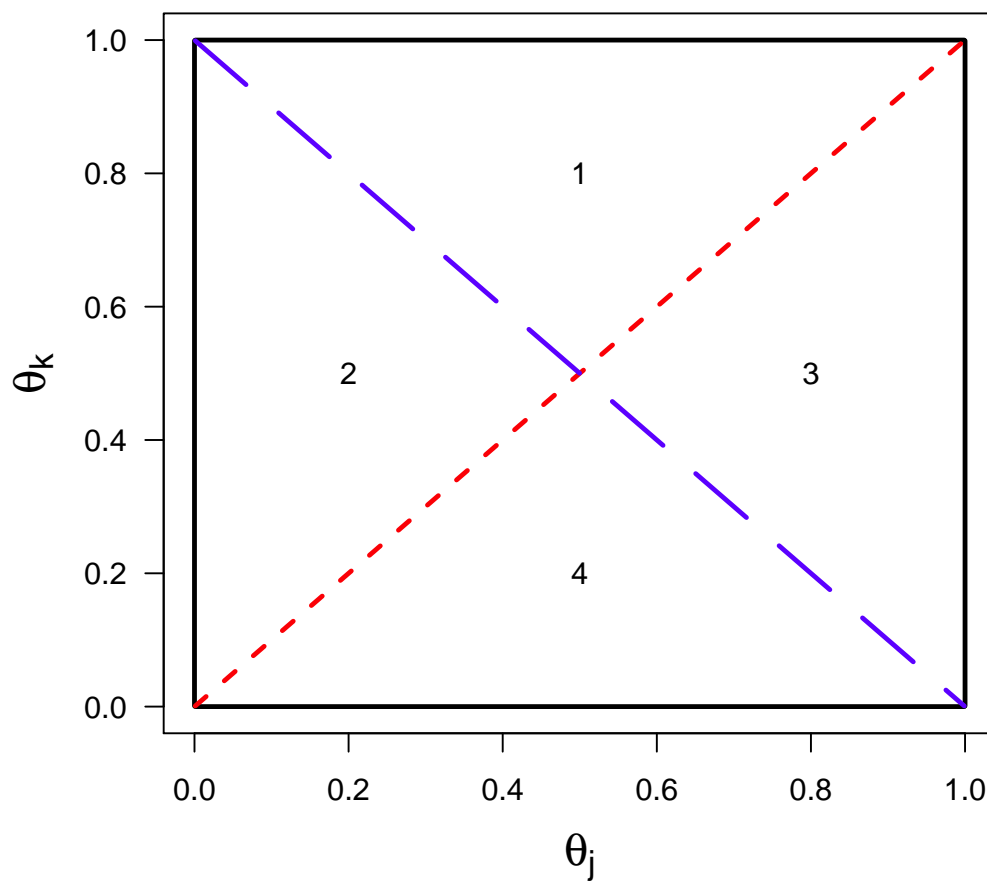


Figure 3.6: A plot of θ_j vs θ_k .

For region 1, points of (θ_j, θ_k) satisfy

$$\begin{aligned}\theta_j &\leq \theta_k, \\ 1 - \theta_j &\leq \theta_k,\end{aligned}$$

which results in

$$1 \leq 2\theta_k \implies \theta_k \geq \frac{1}{2},$$

and this becomes

$$\frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \geq \frac{1}{2},$$

\implies

$$\beta_0 + x_k \beta_1 \geq \log \left(\frac{\frac{1}{2}}{1 - \frac{1}{2}} \right)$$

\implies

$$\boxed{0 \leq \beta_0 + x_k \beta_1.} \tag{3.23}$$

For region 2

$$\begin{aligned}\theta_j &\leq \theta_k, \\ 1 - \theta_j &\geq \theta_k \implies \theta_j - 1 \leq -\theta_k,\end{aligned}$$

which results in

$$2\theta_j - 1 \leq 0 \implies \theta_j \leq \frac{1}{2},$$

and this becomes as

$$\frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \leq \frac{1}{2},$$

\implies

$$\beta_0 + x_j \beta_1 \leq \log \left(\frac{\frac{1}{2}}{1 - \frac{1}{2}} \right)$$

\implies

$$\boxed{\beta_0 + x_j \beta_1 \leq 0.} \quad (3.24)$$

For region 3

$$\begin{aligned} \theta_j &\geq \theta_k \\ 1 - \theta_j &\leq \theta_k \implies \theta_j - 1 \geq -\theta_k \end{aligned}$$

which results in

$$2\theta_j \geq 1 \implies \theta_j \geq \frac{1}{2},$$

and this becomes as

$$\frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \geq \frac{1}{2},$$

\implies

$$\beta_0 + x_j \beta_1 \geq \log \left(\frac{\frac{1}{2}}{1 - \frac{1}{2}} \right)$$

\implies

$$\boxed{0 \leq \beta_0 + x_j \beta_1.} \quad (3.25)$$

For region 4

$$\begin{aligned} \theta_j &\geq \theta_k, \\ 1 - \theta_j &\geq \theta_k, \end{aligned}$$

which results in

$$1 \geq 2\theta_k \implies \theta_k \leq \frac{1}{2},$$

and this becomes as

$$\frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \leq \frac{1}{2},$$

\Rightarrow

$$\beta_0 + x_k \beta_1 \leq \log \left(\frac{\frac{1}{2}}{1 - \frac{1}{2}} \right)$$

\Rightarrow

$$\boxed{\beta_0 + x_k \beta_1 \leq 0.} \quad (3.26)$$

Now, merging (3.23), (3.24), (3.25) and (3.26). Note that from (3.23)

$$0 \leq \beta_0 + x_k \beta_1, \quad (3.27)$$

\Rightarrow

$$\boxed{\frac{-\beta_0}{\beta_1} \leq x_k.} \quad (3.28)$$

The same can be done for (3.24), (3.25) and (3.26), to have

$$\boxed{x_j \leq \frac{-\beta_0}{\beta_1}, \quad \frac{-\beta_0}{\beta_1} \leq x_j \text{ and } x_k \leq \frac{-\beta_0}{\beta_1},} \quad (3.29)$$

respectively.

The relation from θ_j and θ_k to β_0 and β_1 can also be seen from another corner. If

$$a_1 < \theta_j < a_2 \quad \Rightarrow \quad a_1 < \theta_j = \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} < a_2,$$

and

$$b_1 < \theta_k < b_2 \quad \Rightarrow \quad b_1 < \theta_k = \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} < b_2.$$

$$\Rightarrow \quad \log \left(\frac{a_1}{1 - a_1} \right) < \beta_0 + x_j \beta_1 < \log \left(\frac{a_2}{1 - a_2} \right), \quad (3.30)$$

and

$$\log \left(\frac{b_1}{1 - b_1} \right) < \beta_0 + x_k \beta_1 < \log \left(\frac{b_2}{1 - b_2} \right). \quad (3.31)$$

Let

$$\begin{aligned}\lambda_1 &= \log \left(\frac{a_1}{1-a_1} \right) , & \lambda_2 &= \log \left(\frac{a_2}{1-a_2} \right), \\ \lambda_3 &= \log \left(\frac{b_1}{1-b_1} \right) , & \lambda_4 &= \log \left(\frac{b_2}{1-b_2} \right),\end{aligned}$$

then, from (3.30) and (3.31), solve

$$\lambda_1 < \beta_0 + x_j \beta_1 \tag{3.32}$$

$$\text{and} \quad \beta_0 + x_k \beta_1 < \lambda_4, \tag{3.33}$$

The inequalities in (3.32) and (3.33) can be rewritten as

$$-\beta_0 - x_j \beta_1 < -\lambda_1, \tag{3.34}$$

$$\beta_0 + x_k \beta_1 < \lambda_4. \tag{3.35}$$

Then

$$\beta_1 (x_k - x_j) < \lambda_4 - \lambda_1,$$

\implies

$$\boxed{\beta_1 < \frac{\lambda_4 - \lambda_1}{x_k - x_j}}, \tag{3.36}$$

given that $x_j < x_k$, where without loss of generality, the following cases can be considered

$$\left. \begin{aligned} x_j &> 0, \ x_k > 0, \\ x_j &< 0, \ x_k < 0, \\ x_j &< 0, \ x_k > 0. \end{aligned} \right\} \tag{3.37}$$

3.2.1 Positive x_j and x_k

Multiply (3.34) by $\frac{x_k}{x_j}$ to get

$$-\frac{x_k}{x_j} \beta_0 - \frac{x_k}{x_j} x_j \beta_1 < -\frac{x_k}{x_j} \lambda_1,$$

$$\text{and add to (3.35),} \quad \beta_0 + x_k \beta_1 < \lambda_4,$$

to get

$$\beta_0 \left(1 - \frac{x_k}{x_j}\right) < \lambda_4 - \frac{x_k}{x_j} \lambda_1$$

\Rightarrow

$$\boxed{\beta_0 < \frac{\lambda_4 - \frac{x_k}{x_j} \lambda_1}{1 - \frac{x_k}{x_j}}.} \quad (3.38)$$

Furthermore, from (3.30) and (3.31),

$$\beta_0 + x_j \beta_1 < \lambda_2 \quad \Rightarrow \quad -\lambda_2 < -\beta_0 - x_j \beta_1, \quad (3.39)$$

$$\text{and} \quad \lambda_3 < \beta_0 + x_k \beta_1. \quad (3.40)$$

Then

$$\lambda_3 - \lambda_2 < \beta_1 (x_k - x_j)$$

\Rightarrow

$$\boxed{\frac{\lambda_3 - \lambda_2}{x_k - x_j} < \beta_1,} \quad (3.41)$$

and after multiplying (3.39) by $\frac{x_k}{x_j}$, (3.39) and (3.40) are written as

$$-\frac{x_k}{x_j} \lambda_2 < -\frac{x_k}{x_j} \beta_0 - \frac{x_k}{x_j} x_j \beta_1,$$

$$\lambda_3 < \beta_0 + x_k \beta_1,$$

which implies

$$\lambda_3 - \frac{x_k}{x_j} \lambda_2 < \beta_0 \left(1 - \frac{x_k}{x_j}\right)$$

\Rightarrow

$$\boxed{\frac{\lambda_3 - \frac{x_k}{x_j} \lambda_2}{1 - \frac{x_k}{x_j}} < \beta_0.} \quad (3.42)$$

Finally, merging (3.36) with (3.41) yields

$$\boxed{\frac{\lambda_3 - \lambda_2}{x_k - x_j} < \beta_1 < \frac{\lambda_4 - \lambda_1}{x_k - x_j}}, \quad (3.43)$$

and merging (3.38) and (3.42) yields

$$\boxed{\frac{\lambda_3 - \frac{x_k}{x_j}\lambda_2}{1 - \frac{x_k}{x_j}} < \beta_0 < \frac{\lambda_4 - \frac{x_k}{x_j}\lambda_1}{1 - \frac{x_k}{x_j}}}. \quad (3.44)$$

If (3.43) is multiplied by $x_j > 0$ and added to (3.44), then

$$\frac{\lambda_3 - \frac{x_k}{x_j}\lambda_2}{1 - \frac{x_k}{x_j}} + x_j \cdot \frac{\lambda_3 - \lambda_2}{x_k - x_j} < \beta_0 + x_j\beta_1 < \frac{\lambda_4 - \frac{x_k}{x_j}\lambda_1}{1 - \frac{x_k}{x_j}} + x_j \cdot \frac{\lambda_4 - \lambda_1}{x_k - x_j}, \quad (3.45)$$

and if (3.43) is multiplied by $x_k > 0$ and added to (3.44), then

$$\frac{\lambda_3 - \frac{x_k}{x_j}\lambda_2}{1 - \frac{x_k}{x_j}} + x_k \cdot \frac{\lambda_3 - \lambda_2}{x_k - x_j} < \beta_0 + x_k\beta_1 < \frac{\lambda_4 - \frac{x_k}{x_j}\lambda_1}{1 - \frac{x_k}{x_j}} + x_k \cdot \frac{\lambda_4 - \lambda_1}{x_k - x_j}, \quad (3.46)$$

where both of (3.45) and (3.46) can be added and divided by 2 to get

$$\boxed{\begin{aligned} \frac{\lambda_3 - \frac{x_k}{x_j}\lambda_2}{1 - \frac{x_k}{x_j}} + \frac{(x_j + x_k)}{2} \cdot \frac{\lambda_3 - \lambda_2}{x_k - x_j} &< \beta_0 + \frac{(x_j + x_k)}{2}\beta_1 < \\ \frac{\lambda_4 - \frac{x_k}{x_j}\lambda_1}{1 - \frac{x_k}{x_j}} + \frac{(x_j + x_k)}{2} \cdot \frac{\lambda_4 - \lambda_1}{x_k - x_j}. \end{aligned}} \quad (3.47)$$

3.2.2 Negative x_j and x_k

Inequality (3.47) changes if $x_j < 0$ and $x_k < 0$. If (3.43) is multiplied by $x_j < 0$ and (3.44) by -1, then adding them gives

$$-\left[\frac{\lambda_4 - \frac{x_k}{x_j}\lambda_1}{1 - \frac{x_k}{x_j}}\right] + x_j \cdot \frac{\lambda_4 - \lambda_1}{x_k - x_j} < x_j\beta_1 - \beta_0 < -\left[\frac{\lambda_3 - \frac{x_k}{x_j}\lambda_2}{1 - \frac{x_k}{x_j}}\right] + x_j \cdot \frac{\lambda_3 - \lambda_2}{x_k - x_j}, \quad (3.48)$$

and if (3.43) is multiplied by $x_k < 0$ and (3.44) by -1, then adding them gives

$$-\left[\frac{\lambda_4 - \frac{x_k}{x_j}\lambda_1}{1 - \frac{x_k}{x_j}}\right] + x_k \cdot \frac{\lambda_4 - \lambda_1}{x_k - x_j} < x_k\beta_1 - \beta_0 < -\left[\frac{\lambda_3 - \frac{x_k}{x_j}\lambda_2}{1 - \frac{x_k}{x_j}}\right] + x_k \cdot \frac{\lambda_3 - \lambda_2}{x_k - x_j}, \quad (3.49)$$

where both of (3.48) and (3.49) can be added and divided by 2 to get

$$\boxed{-\left[\frac{\lambda_4 - \frac{x_k}{x_j}\lambda_1}{1 - \frac{x_k}{x_j}}\right] + \frac{(x_j + x_k)}{2} \cdot \frac{\lambda_4 - \lambda_1}{x_k - x_j} < -\beta_0 + \frac{(x_j + x_k)}{2}\beta_1 < -\left[\frac{\lambda_3 - \frac{x_k}{x_j}\lambda_2}{1 - \frac{x_k}{x_j}}\right] + \frac{(x_j + x_k)}{2} \cdot \frac{\lambda_3 - \lambda_2}{x_k - x_j}.} \quad (3.50)$$

Inequalities (3.43), (3.44), (3.47) and (3.50) determine the set of β points that correspond to θ (in case of $x_j > 0, x_k > 0$ or $x_j < 0, x_k < 0$). This can be better seen by looking at figures 3.7, 3.8 and 3.9.

3.2.3 Negative x_j and Positive x_k

If $x_j < 0$ and $x_k > 0$, then multiplying (3.32) by $\frac{x_k}{x_j}$ gives

$$\frac{x_k}{x_j}\beta_0 + \frac{x_k}{x_j}x_j\beta_1 < \frac{x_k}{x_j}\lambda_1$$

$$\text{and add to (3.35),} \quad \beta_0 + x_k\beta_1 < \lambda_4,$$

to get

$$\beta_0 \left(1 + \frac{x_k}{x_j}\right) + \beta_1 (2x_k) < \lambda_4 + \frac{x_k}{x_j}\lambda_1. \quad (3.51)$$

Also, after multiplying (3.39) by $\frac{x_k}{x_j}$,

$$\frac{x_k}{x_j}\lambda_2 < \frac{x_k}{x_j}\beta_0 + \frac{x_k}{x_j}x_j\beta_1,$$

$$\text{and add to (3.40),} \quad \lambda_3 < \beta_0 + x_k\beta_1,$$

to get

$$\lambda_3 + \frac{x_k}{x_j} \lambda_2 < \beta_0 \left(1 + \frac{x_k}{x_j} \right) + \beta_1 (2x_k). \quad (3.52)$$

Inequalities (3.51) and (3.52) are merged as

$$\boxed{\lambda_3 + \frac{x_k}{x_j} \lambda_2 < \left(1 + \frac{x_k}{x_j} \right) \beta_0 + 2x_k \beta_1 < \lambda_4 + \frac{x_k}{x_j} \lambda_1.} \quad (3.53)$$

Plots of maps from intervals of $\boldsymbol{\theta} = [\theta_j, \theta_k]'$ to intervals of $\boldsymbol{\beta} = [\beta_0, \beta_1]'$ are shown in figures 3.7, 3.8 and 3.9 where rectangles in $\boldsymbol{\theta}$ transmit to convex parallelograms in $\boldsymbol{\beta}$. Every $\boldsymbol{\beta}$ point has to fit inequality (3.53) except in the bottom-right plot of Figure 3.8 which fits (3.50) because x_j and x_k are fixed at negative values. The same thing is for bottom-right plot of Figure 3.9 which fits (3.47) because x_j and x_k are fixed at positive values.

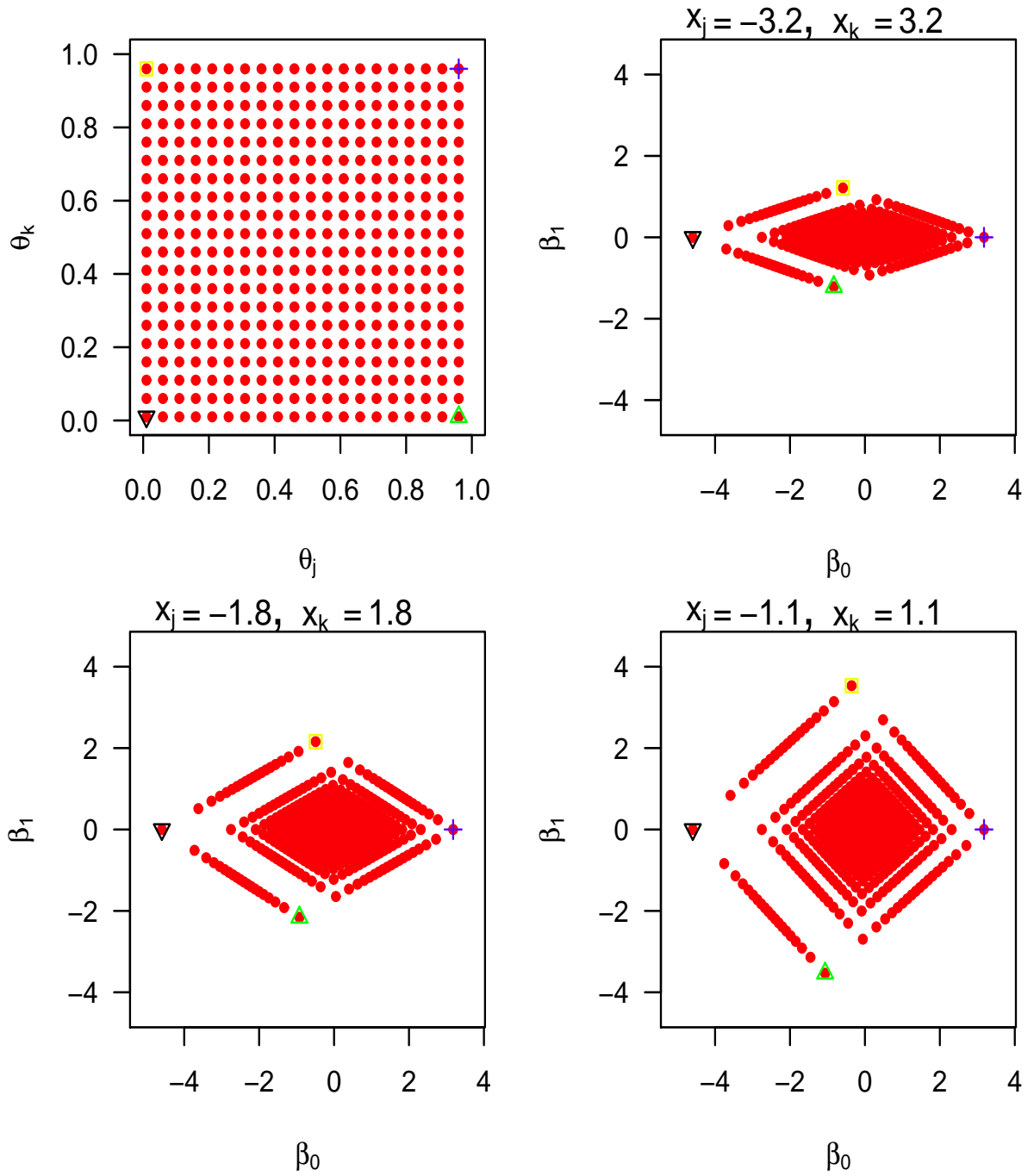


Figure 3.7: Transformation plots from θ_j, θ_k to β_0, β_1 under symmetric fixed x_j and x_k . top-right: A plot of β_0 vs β_1 that corresponds to combinations of $\theta_j \in [0.05, 0.95]$ and $\theta_k \in [0.05, 0.95]$ with $x_j = -3.2$ and $x_k = 3.2$. Bottom-left: $x_j = -1.8$ and $x_k = 1.8$. Bottom-right: $x_j = -1.1$ and $x_k = 1.1$.

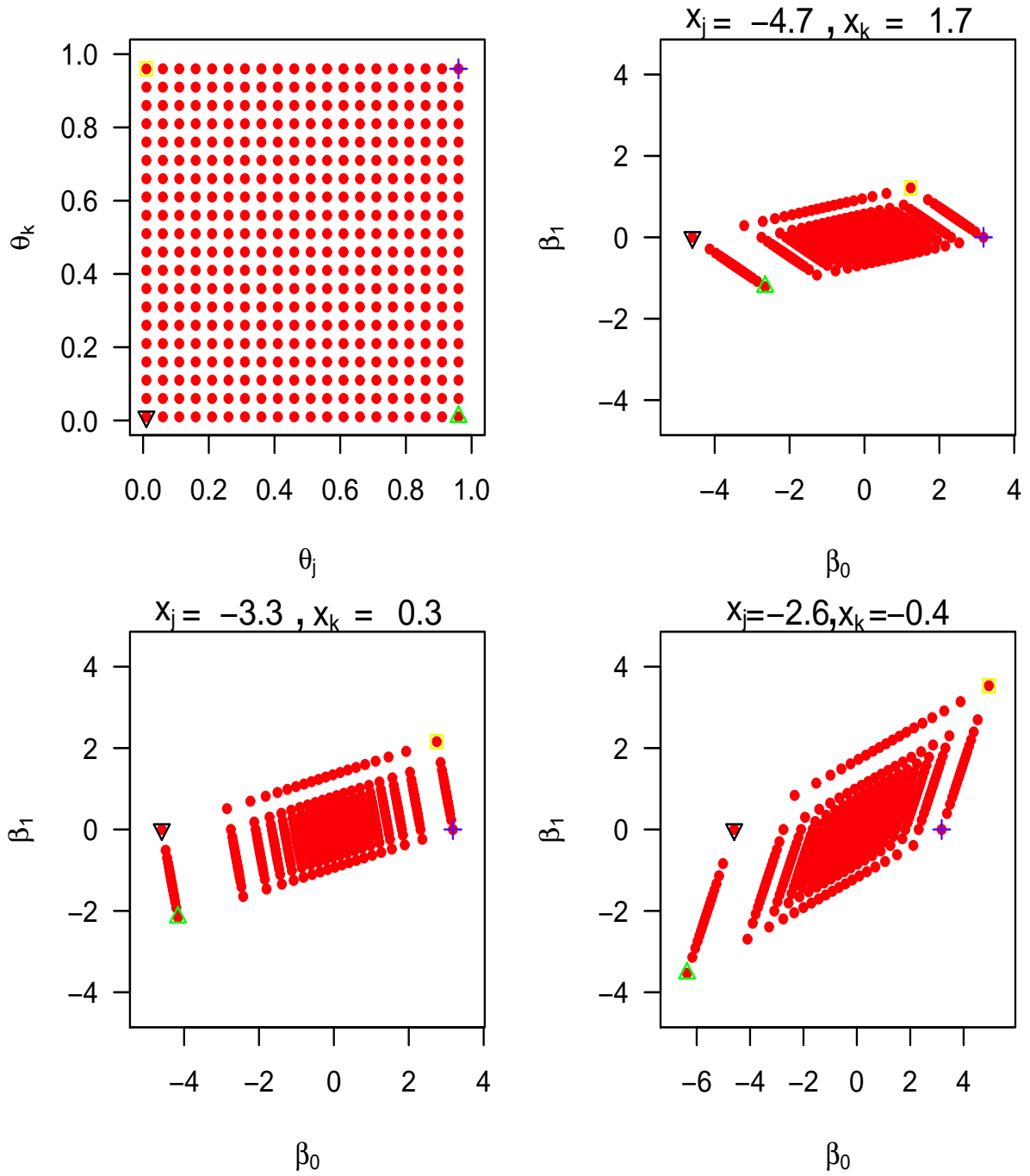


Figure 3.8: Transformation plots from θ_j, θ_k to β_0, β_1 under left-shifted fixed x_j and x_k . top-right: A plot of β_0 vs β_1 that corresponds to combinations of $\theta_j \in [0.05, 0.95]$ and $\theta_k \in [0.05, 0.95]$ with $x_j = -4.7$ and $x_k = 1.7$. Bottom-left: $x_j = -3.3$ and $x_k = 0.3$. Bottom-right: $x_j = -2.6$ and $x_k = -0.4$.

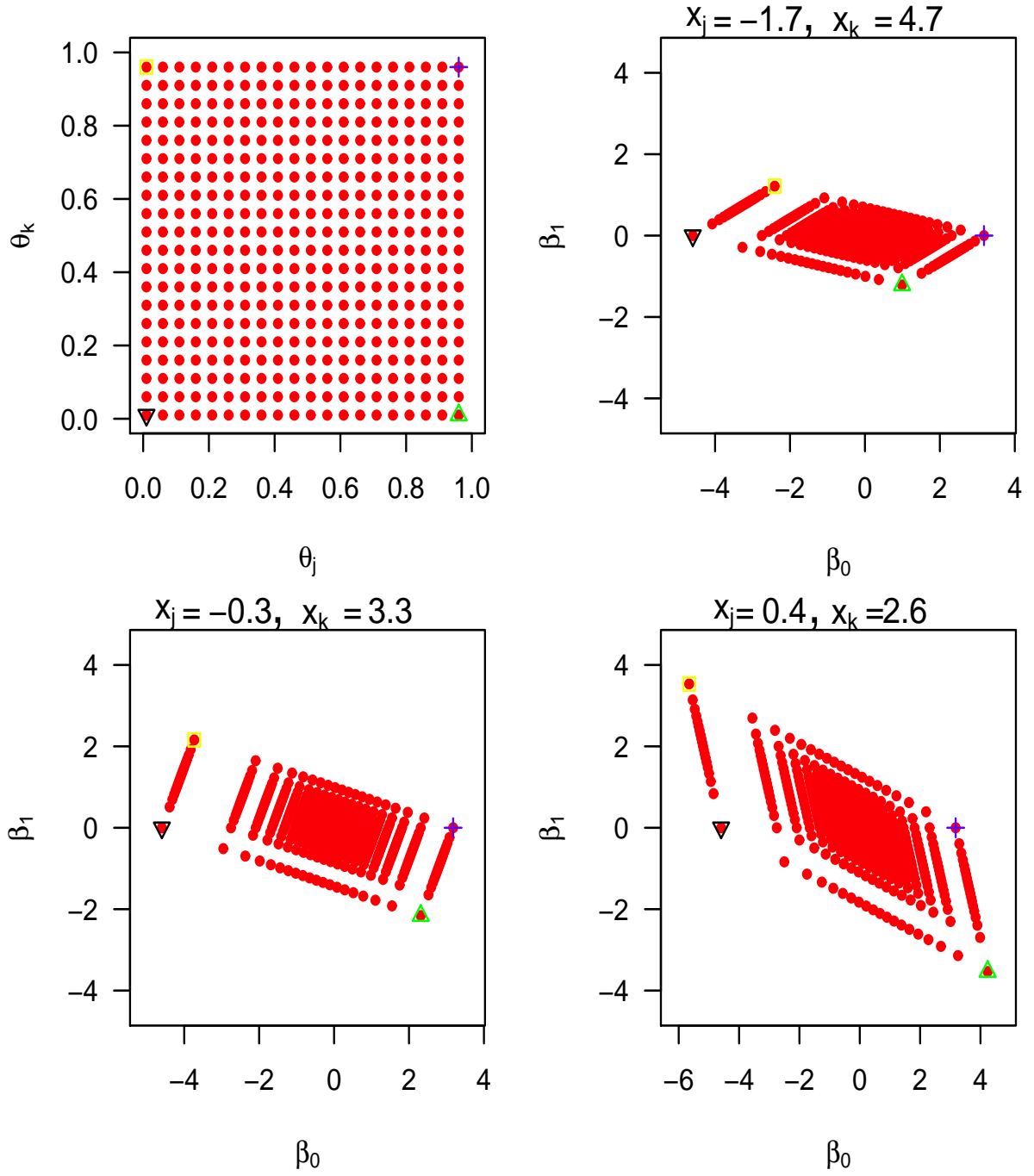


Figure 3.9: Transformation plots from θ_j, θ_k to β_0, β_1 under right-shifted fixed x_j and x_k . top-right: A plot of β_0 vs β_1 that corresponds to combinations of $\theta_j \in [0.05, 0.95]$ and $\theta_k \in [0.05, 0.95]$ with $x_j = -1.7$ and $x_k = 4.7$. Bottom-left: $x_j = -0.3$ and $x_k = 3.3$. Bottom-right: $x_j = 0.4$ and $x_k = 2.6$.

3.3 2-Parameter Increasing ImpLogit Model

This section builds on Section 2.1 is to involve imprecision in the increasing ImpLogit model. In this case, β_1 interval estimate has to have positive limits but no restrictions on interval limits of the regression parameter β_0 . Let $H_j = \varphi_{jk1}$, $H_k - H_j = \varphi_{jk2}$ and $1 - H_k = \varphi_{jk3}$ where $\varphi_{jk1} + \varphi_{jk2} + \varphi_{jk3} = 1$. Then the posterior distribution given in (2.12) is rewritten, under a selected θ_j, θ_k , as

$$\begin{aligned} \pi(\boldsymbol{\beta}|y_1, \dots, y_m) \propto f(\mathbf{y}, \boldsymbol{\beta}) &= \left\{ \prod_{i=1}^m \left[\binom{n_i}{y_i} \left(\frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right] \right\} \\ &\times \left[\frac{\Gamma(\nu)}{\Gamma(\nu \varphi_{jk1}) \Gamma(\nu \varphi_{jk2}) \Gamma(\nu \varphi_{jk3})} \right] \\ &\times \left[\left(\frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right)^{\nu \varphi_{jk1} - 1} \left(\frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} - \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right)^{\nu \varphi_{jk2} - 1} \right. \\ &\quad \left. \cdot \left(1 - \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right)^{\nu \varphi_{jk3} - 1} \right] \\ &\times |x_k - x_j| \left\{ \left[\frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \right] \left[\frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right] \right\}, \end{aligned} \quad (3.54)$$

where the prior parameters sum to

$$\begin{aligned} \nu H_j + (\nu H_k - \nu H_j) + (\nu - \nu H_k) &= \nu H_j + \nu (H_k - H_j) + \nu (1 - H_k) \\ &= \nu \varphi_{jk1} + \nu \varphi_{jk2} + \nu \varphi_{jk3} = \nu (\varphi_{jk1} + \varphi_{jk2} + \varphi_{jk3}) = \nu. \end{aligned}$$

Also, since $H_j \in [0, 1]$ then

$$\min(\varphi_{jk1}) = \min(H_j) = 0 \quad , \quad \max(\varphi_{jk1}) = \max(H_j) = 1,$$

and

$$\min(\varphi_{jk3}) = \min(1 - H_k) = 0 \quad , \quad \max(\varphi_{jk3}) = \max(1 - H_k) = 1.$$

If

$$\varphi_{jk1} \longrightarrow 0 \quad \text{and} \quad \varphi_{jk3} \longrightarrow 1, \quad \text{then} \quad \varphi_{jk2} \longrightarrow 0, \quad (3.55)$$

and

$$\varphi_{jk1} \longrightarrow 1 \quad \text{and} \quad \varphi_{jk3} \longrightarrow 0, \quad \text{then} \quad \varphi_{jk2} \longrightarrow 0, \quad (3.56)$$

The prior distribution in (2.11) is not considered as a pdf when $\nu = 0$, but it is called an improper prior. This is because if $\nu = 0$ then $\Gamma(\nu)$ in (2.11) is undefined. The posterior distribution in (3.54) becomes an improper one if $\nu = 0$ and $y_i = 0$ or n_i , $\forall i$. The same thing happens if $\varphi_{jkl} = 0$ or 1 , $\forall l = 1, 2, 3$, and $y_i = 0$ or n_i , $\forall i$. In these cases, the improper prior does not define a posterior distribution.

The joint posterior densities in (3.54) can now be fitted, subject to (3.55) and (3.56), to find the interval estimates of β_0 and β_1 . Since all φ_{jk1} , φ_{jk2} and φ_{jk3} belong to $[0, 1]$, then a set of logistic function fits can be performed for (3.54) with a set of values of φ_{jk1} , φ_{jk2} and φ_{jk3} . Estimating regression parameters in (3.54) at cases of (3.55) and (3.56) does not necessarily produce lower and upper estimates. This situation is different from that in Section 1.7 where a re-look at (1.60) and (1.61) shows that lower and upper expectations correspond to limiting values of φ_i , $\forall i$. The induced prior pdf in (2.11) is not a conjugate one to the posterior density function in (3.54). Consequently, estimating the regression parameters under (3.55) and (3.56) does not suffice to find their interval estimates, but should be done under a set of values of φ_{jk1} , φ_{jk2} and φ_{jk3} .

Tables 3.1, 3.2 and 3.3 provide the computed lower, upper and interval estimates for β_0 and β_1 when $\nu = 0.5, 1$ and 2 , respectively, for the real data given in Chapter 2 in Table 2.1. Computations are performed by following Metropolis Hastings algorithm. The class of logistic fits is gained by changing the values of φ_{jk1} , φ_{jk2} and φ_{jk3} , provided that they sum to 1 and $H_j < H_k$, $\forall j, k$. Then $H_j = \varphi_{jk1}$ was assigned 19 values and other 19 values are given to $H_k = 1 - \varphi_{jk3}$. The assigned values for H_j are

$$0.001, 0.14, 0.19, \dots, 0.998, \quad (3.57)$$

while the values for H_k are

$$0.002, 0.15, 0.20, \dots, 0.999, \quad (3.58)$$

provided that $H_j < H_k$, $\forall j, k$. This means that there is $19+18+\dots+1 = 190$ logistic functions to be fitted in each ImpLogit fit.

Table 3.1: Interval estimates of β_0 and β_1 that correspond to top plots in Figure 3.10 for 2-parameter increasing ImpLogit model when $\nu = 0.5$. $\text{imp}(\hat{\beta}_i)$ refers to the interval estimate length.

selected pair	$\hat{\beta}_0$	$\bar{\hat{\beta}}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\hat{\beta}}_1$	$\text{imp}(\hat{\beta}_1)$
$\theta_1, \theta_6(x_1, x_6)$	-3.226	-3.221	0.005	0.589	0.613	0.024
$\theta_2, \theta_5(x_2, x_5)$	-3.227	-3.222	0.005	0.586	0.612	0.026
$\theta_3, \theta_4(x_3, x_4)$	-3.228	-3.223	0.005	0.585	0.613	0.028

Table 3.2: Interval estimates of β_0 and β_1 that correspond to middle plots in Figure 3.10 for 2-parameter increasing ImpLogit model when $\nu = 1$. $\text{imp}(\hat{\beta}_i)$ refers to the interval estimate length.

selected pair	$\hat{\beta}_0$	$\bar{\hat{\beta}}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\hat{\beta}}_1$	$\text{imp}(\hat{\beta}_1)$
$\theta_1, \theta_6(x_1, x_6)$	-3.226	-3.221	0.005	0.595	0.621	0.026
$\theta_2, \theta_5(x_2, x_5)$	-3.227	-3.218	0.009	0.593	0.622	0.029
$\theta_3, \theta_4(x_3, x_4)$	-3.228	-3.216	0.012	0.591	0.624	0.033

Table 3.3: Interval estimates of β_0 and β_1 that correspond to bottom plots in Figure 3.10 for 2-parameter increasing ImpLogit model when $\nu = 2$. $\text{imp}(\hat{\beta}_i)$ refers to the interval estimate length.

selected pair	$\hat{\beta}_0$	$\bar{\hat{\beta}}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\hat{\beta}}_1$	$\text{imp}(\hat{\beta}_1)$
$\theta_1, \theta_6(x_1, x_6)$	-3.227	-3.219	0.008	0.597	0.624	0.027
$\theta_2, \theta_5(x_2, x_5)$	-3.229	-3.216	0.013	0.594	0.624	0.030
$\theta_3, \theta_4(x_3, x_4)$	-3.229	-3.215	0.014	0.592	0.626	0.034

Numerical results show that increasing ν value results in having higher imprecision (longer interval estimates) in regression parameters. Having a closer sequential order of selected θ_j and θ_k increases the lengths of interval estimates of β_0 and β_1 .

Tables 3.4, 3.5 and 3.6 present interval estimates of θ when $x = 5$ that correspond to regression parameters interval estimates in tables 3.1, 3.2 and 3.3, respectively. Since $x_i \geq 0, \forall i = 1, \dots, 6$, for data given in Table 2.1, then based on inequality (3.16), each $\hat{\theta}$ corresponds to $\underline{\beta} = [\hat{\beta}_0, \hat{\beta}_1]'$ and each $\bar{\theta}$ corresponds to $\bar{\beta} = [\bar{\beta}_0, \bar{\beta}_1]'$.

ImpLogit plots are given in Figure 3.10. Each ImpLogit plot in Figure 3.10 is a collection of 190 logistic function fits.

Table 3.4: Interval estimates of θ for Table 3.1 when $x = 5$.

selected pair	$\hat{\theta}$	$\bar{\theta}$	$\text{imp}(\hat{\theta})$
$\theta_1, \theta_6(x_1, x_6)$	0.430	0.461	0.031
$\theta_2, \theta_5(x_2, x_5)$	0.426	0.460	0.034
$\theta_3, \theta_4(x_3, x_4)$	0.425	0.461	0.036

Table 3.5: Interval estimates of θ for Table 3.2 when $x = 5$.

selected pair	$\hat{\theta}$	$\bar{\theta}$	$\text{imp}(\hat{\theta})$
$\theta_1, \theta_6(x_1, x_6)$	0.438	0.471	0.033
$\theta_2, \theta_5(x_2, x_5)$	0.435	0.473	0.038
$\theta_3, \theta_4(x_3, x_4)$	0.432	0.476	0.044

Table 3.6: Interval estimates of θ for Table 3.3 when $x = 5$.

selected pair	$\hat{\theta}$	$\bar{\theta}$	$\text{imp}(\hat{\theta})$
$\theta_1, \theta_6(x_1, x_6)$	0.440	0.475	0.035
$\theta_2, \theta_5(x_2, x_5)$	0.436	0.476	0.040
$\theta_3, \theta_4(x_3, x_4)$	0.433	0.479	0.046

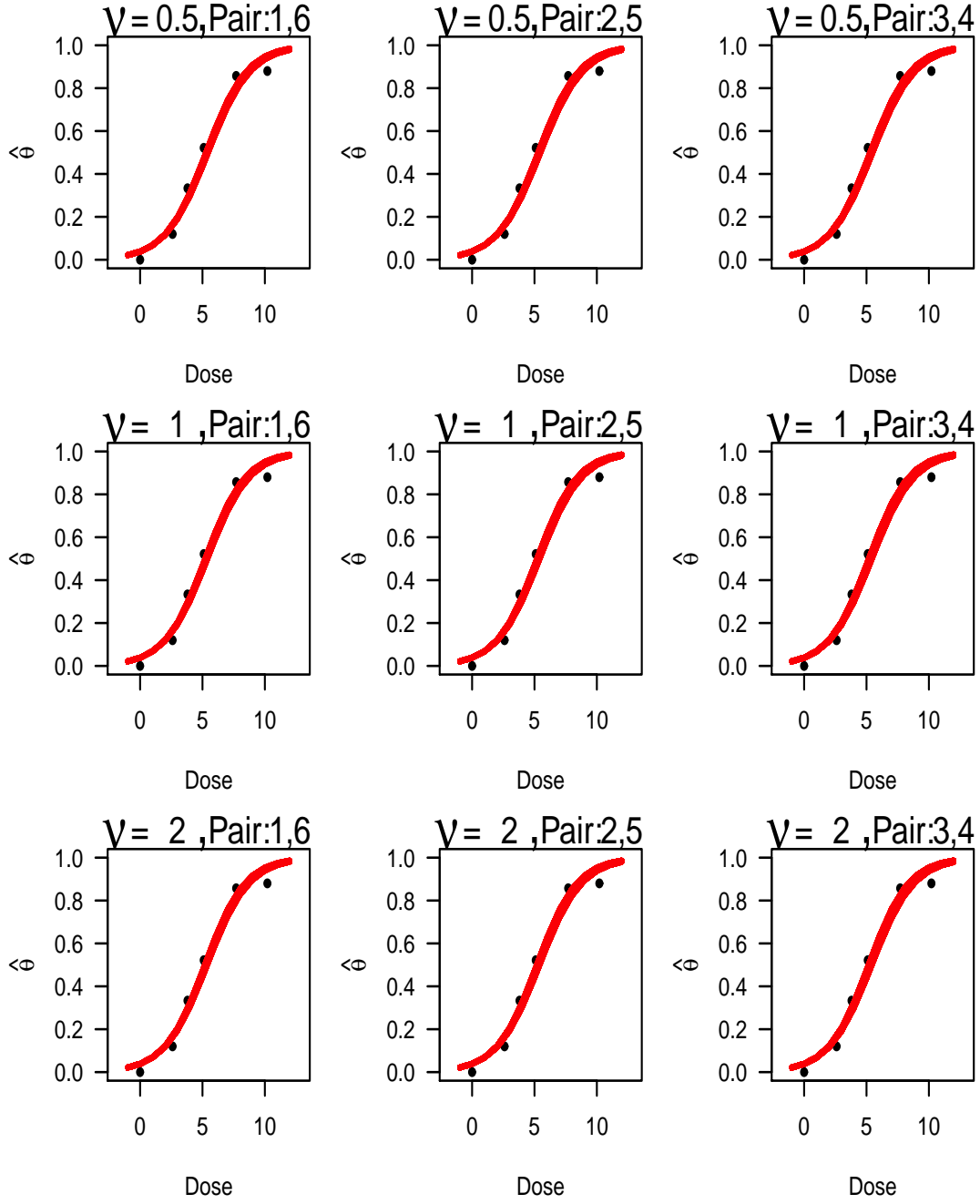


Figure 3.10: Top: Plots of increasing ImpLogit model for the real data given in Table 2.1 with $\nu = 0.5$ and selected pairs θ_1, θ_6 (top-left), θ_2, θ_5 (top-middle) and θ_3, θ_4 (top-right). Middle: $\nu = 1$ and selected pairs θ_1, θ_6 (middle-left), θ_2, θ_5 (middle-middle) and θ_3, θ_4 (middle-right). Bottom: $\nu = 2$ and selected pairs θ_1, θ_6 (bottom-left), θ_2, θ_5 (bottom-middle) and θ_3, θ_4 (bottom-right).

3.4 2-Parameter Decreasing ImpLogit Model

This section presents ImpLogit model with negative interval limits for β_1 . Let $\varphi_{jk1}^* = 1 - H_j^*$, $\varphi_{jk2}^* = H_j^* - H_k^*$ and $\varphi_{jk3}^* = H_k^*$, then the posterior distribution in (2.38) is rewritten as

$$\begin{aligned}
\pi(\boldsymbol{\beta}|y_1, \dots, y_m) &\propto f(\mathbf{y}, \boldsymbol{\beta}) \\
&= \left\{ \prod_{i=1}^m \left[\binom{n_i}{y_i} \left(\frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right] \right\} \\
&\times \left[\frac{\Gamma(\nu)}{\Gamma(\nu \varphi_{jk1}^*) \Gamma(\nu \varphi_{jk2}^*) \Gamma(\nu \varphi_{jk3}^*)} \right] \\
&\times \left[\left(1 - \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right)^{\nu \varphi_{jk1}^* - 1} \right. \\
&\quad \cdot \left. \left(\frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} - \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right)^{\nu \varphi_{jk2}^* - 1} \left(\frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right)^{\nu \varphi_{jk3}^*} \right] \\
&\times |x_k - x_j| \left\{ \left[\frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \right] \left[\frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right] \right\} \tag{3.59}
\end{aligned}$$

Now, the regression parameters in (3.59) can be estimated. Tables 3.7, 3.8 and 3.9 confirm that β_1 interval estimates include only negative values which fits the decreasing ImpLogit model for data given in Table 2.4. The assigned values for H_j^* are

$$0.999, \dots, 0.20, 0.15, 0.002, \tag{3.60}$$

while the values for H_k^* are

$$0.998, \dots, 0.19, 0.14, 0.001, \tag{3.61}$$

provided that $H_j^* > H_k^*$, $\forall j, k$. Increasing ν value results in having longer interval estimates of regression parameters.

Table 3.7: Interval estimates of β_0 and β_1 that correspond to top plots in Figure 3.11 for 2-parameter decreasing ImpLogit model when $\nu = 0.5$.

selected pair	$\hat{\beta}_0$	$\bar{\hat{\beta}}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\hat{\beta}}_1$	$\text{imp}(\hat{\beta}_1)$
$\theta_1, \theta_6(x_1, x_6)$	3.221	3.227	0.006	-0.612	-0.593	0.019
$\theta_2, \theta_5(x_2, x_5)$	3.220	3.228	0.008	-0.614	-0.591	0.023
$\theta_3, \theta_4(x_3, x_4)$	3.218	3.229	0.011	-0.616	-0.588	0.028

Table 3.8: Interval estimates of β_0 and β_1 that correspond to middle plots in Figure 3.11 for 2-parameter decreasing ImpLogit model when $\nu = 1$.

selected pair	$\hat{\beta}_0$	$\bar{\hat{\beta}}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\hat{\beta}}_1$	$\text{imp}(\hat{\beta}_1)$
$\theta_1, \theta_6(x_1, x_6)$	3.220	3.228	0.008	-0.614	-0.591	0.023
$\theta_2, \theta_5(x_2, x_5)$	3.219	3.229	0.010	-0.616	-0.589	0.027
$\theta_3, \theta_4(x_3, x_4)$	3.217	3.231	0.014	-0.618	-0.586	0.032

Table 3.9: Interval estimates of β_0 and β_1 that correspond to bottom plots in Figure 3.11 for 2-parameter decreasing ImpLogit model when $\nu = 2$.

selected pair	$\hat{\beta}_0$	$\bar{\hat{\beta}}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\hat{\beta}}_1$	$\text{imp}(\hat{\beta}_1)$
$\theta_1, \theta_6(x_1, x_6)$	3.218	3.232	0.014	-0.615	-0.590	0.025
$\theta_2, \theta_5(x_2, x_5)$	3.216	3.233	0.017	-0.617	-0.585	0.032
$\theta_3, \theta_4(x_3, x_4)$	3.215	3.231	0.016	-0.618	-0.584	0.034

Interval estimates of θ are in tables 3.10, 3.11 and 3.12. Figure 3.11 gives the decreasing ImpLogit fits.

Table 3.10: Interval estimates of θ for Table 3.7 when $x = 5$.

selected pair	$\hat{\theta}$	$\bar{\hat{\theta}}$	$\text{imp}(\hat{\theta})$
$\theta_1, \theta_6(x_1, x_6)$	0.540	0.565	0.025
$\theta_2, \theta_5(x_2, x_5)$	0.537	0.568	0.031
$\theta_3, \theta_4(x_3, x_4)$	0.534	0.572	0.038

Table 3.11: Interval estimates of θ for Table 3.8 when $x = 5$.

selected pair	$\hat{\theta}$	$\bar{\hat{\theta}}$	$\text{imp}(\hat{\theta})$
$\theta_1, \theta_6(x_1, x_6)$	0.537	0.568	0.031
$\theta_2, \theta_5(x_2, x_5)$	0.535	0.571	0.036
$\theta_3, \theta_4(x_3, x_4)$	0.532	0.575	0.043

Table 3.12: Interval estimates of θ for Table 3.9 when $x = 5$.

selected pair	$\hat{\theta}$	$\bar{\hat{\theta}}$	$\text{imp}(\hat{\theta})$
$\theta_1, \theta_6(x_1, x_6)$	0.536	0.570	0.034
$\theta_2, \theta_5(x_2, x_5)$	0.533	0.576	0.043
$\theta_3, \theta_4(x_3, x_4)$	0.531	0.577	0.046

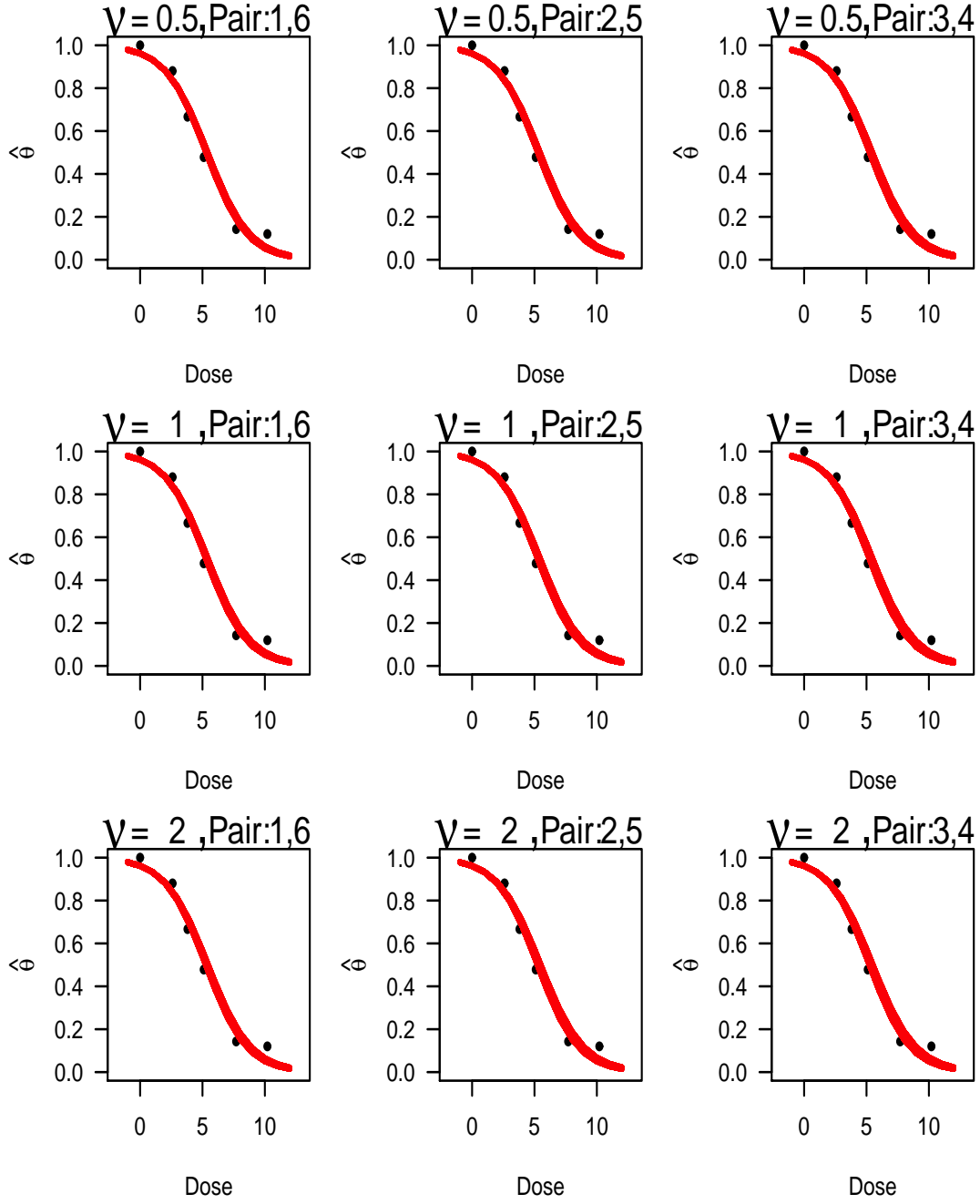


Figure 3.11: Top: Plots of decreasing ImpLogit model for the real data given in Table 2.4 with $\nu = 0.5$ and selected pairs θ_1, θ_6 (top-left), θ_2, θ_5 (top-middle) and θ_3, θ_4 (top-right). Middle: $\nu = 1$ and selected pairs θ_1, θ_6 (middle-left), θ_2, θ_5 (middle-middle) and θ_3, θ_4 (middle-right). Bottom: $\nu = 2$ and selected pairs θ_1, θ_6 (bottom-left), θ_2, θ_5 (bottom-middle) and θ_3, θ_4 (bottom-right).

3.5 Mixture of Beliefs in 2-Parameter ImpLogit Model

Cases of increasing and decreasing ImpLogit models can be generalized and merged in one general form. This happens if there is a mixture of beliefs about the behaviour of the logistic models. To think of the logit model as a mixture of two models, assume that $-\infty < \beta_1 < \infty$, then the joint prior pdf is written as a mixture of two joint pdf's of the increasing and decreasing logistic regression models, $f_1(\theta_j, \theta_k)$ and $f_2(\theta_j, \theta_k)$, as follows

$$f(\theta_j, \theta_k) = w f_1(\theta_j, \theta_k) + (1 - w) f_2(\theta_j, \theta_k), \quad (3.62)$$

where $w \in [0, 1]$ refers to the weight that reflects the degree of belief in $f_1(\theta_j, \theta_k)$ and $f_2(\theta_j, \theta_k)$. If $w = 1$, then $f(\theta_j, \theta_k)$ is the same as in Section 3.3, and if $w = 0$, then the mixture joint density function is similar to that in Section 3.4.

The joint prior pdf is

$$\begin{aligned} \pi(\boldsymbol{\beta} | y_1, \dots, y_m) \propto f(\mathbf{y}, \boldsymbol{\beta}) &= \left\{ \prod_{i=1}^m \left[\binom{n_i}{y_i} \left(\frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right] \right\} \\ &\times w \left[\frac{\Gamma(\nu)}{\Gamma(\nu H_j) \Gamma(\nu H_k - \nu H_j) \Gamma(\nu - \nu H_k)} \right] \\ &\times \left[\left(\frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right)^{\nu H_j - 1} \left(\frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} - \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right)^{\nu H_k - \nu H_j - 1} \right. \\ &\times \left. \left(1 - \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right)^{\nu - \nu H_k - 1} \right] \cdot |x_k - x_j| \left\{ \left[\frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \right] \left[\frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right] \right\} \\ &+ \left\{ \prod_{i=1}^m \left[\binom{n_i}{y_i} \left(\frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right] \right\} \\ &\times (1 - w) \left[\frac{\Gamma(\nu)}{\Gamma(\nu - \nu H_j^*) \Gamma(\nu H_j^* - \nu H_k^*) \Gamma(\nu H_k^*)} \right] \\ &\times \left[\left(1 - \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right)^{\nu - \nu H_j^* - 1} \left(\frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} - \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right)^{\nu H_j^* - \nu H_k^* - 1} \right. \\ &\times \left. \left(\frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right)^{\nu H_k^* - 1} \right] \cdot |x_k - x_j| \left\{ \left[\frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \right] \left[\frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right] \right\}. \quad (3.63) \end{aligned}$$

The correlation between θ_j and θ_k is found by deriving the prior covariance of θ_j and

θ_k from the joint pdf given in (3.62) as

$$\begin{aligned}
\text{Cov}(\theta_j, \theta_k) &= E\{[\theta_j - E(\theta_j)][\theta_k - E(\theta_k)]\} \\
&= wE\{[\theta_j - E_1(\theta_j)][\theta_k - E_1(\theta_k)]\} + (1-w)E\{[\theta_j - E_2(\theta_j)][\theta_k - E_2(\theta_k)]\} \\
&= wE\{[\theta_j - E_1(\theta_j)][\theta_k - E_1(\theta_k)]\} \\
&\quad + (1-w)E\{[\theta_j - E_2(\theta_j)][\theta_k - E_2(\theta_k)]\} \\
&= w\text{Cov}_1(\theta_j, \theta_k) + (1-w)\text{Cov}_2(\theta_j, \theta_k) \\
&= w\frac{H_j(1-H_k)}{\nu+1} + (1-w)\frac{H_k^*(1-H_j^*)}{\nu+1},
\end{aligned} \tag{3.64}$$

where $\text{Cov}_1(\theta_j, \theta_k)$ and $\text{Cov}_2(\theta_j, \theta_k)$ are given in (2.20) and (2.36), respectively. If $w = 1$, then (3.64) becomes as in (2.20), and if $w = 0$, the covariance is similar to that in (2.36). Now, $\text{Var}(\theta_j)$ and $\text{Var}(\theta_k)$ are required. Using the aggregation property gives the probability density function of θ_i as

$$f(\theta_i) = wf_1(\theta_i) + (1-w)f_2(\theta_i). \tag{3.65}$$

Using variance equations in (2.22) and (2.34) gives, for any $i = 1, \dots, m$,

$$\text{Var}(\theta_i) = w\frac{H_i(1-H_i)}{\nu+1} + (1-w)\frac{H_i^*(1-H_i^*)}{\nu+1}. \tag{3.66}$$

Now based on (3.64) and (3.66), the correlation is

$$\begin{aligned}
\text{corr}(\theta_j, \theta_k) &= \\
&\frac{w\frac{H_j(1-H_k)}{\nu+1} + (1-w)\frac{H_k^*(1-H_j^*)}{\nu+1}}{\sqrt{\left[w\frac{H_j(1-H_j)}{\nu+1} + (1-w)\frac{H_j^*(1-H_j^*)}{\nu+1}\right]} \cdot \sqrt{\left[w\frac{H_k(1-H_k)}{\nu+1} + (1-w)\frac{H_k^*(1-H_k^*)}{\nu+1}\right]}}.
\end{aligned} \tag{3.67}$$

If $w = 1$, then (3.67) becomes as in (2.23), and if $w = 0$, the correlation is similar to (2.37). Calculations and graphs of this section and sections 3.6 and 3.7 are given in Appendix D. Tables D.1, D.2 and D.3 provide interval estimates of regression parameters by fitting data in Table 2.1 for different values of w . Interval estimates of θ are in tables D.4, D.5 and D.6. Values of H_j , H_k , H_j^* and H_k^* are the same as in (3.57), (3.58), (3.60) and (3.61).

Increasing w value makes the mixed-belief 2-parameter ImpLogit model closer to be an increasing one, where interval estimates lengths become shorter as can be seen in tables D.1, D.2 and D.3. The same thing can be found in interval estimates of θ at $x = 5$, in tables D.4, D.5 and D.6. ImpLogit fits are in Figure D.1.

3.6 ImpLogit with Imprecise Bivariate Logit-Normal Model

Sections 2.3 and 2.4 are now extended to involve imprecision in logit-normal distribution with application to ImpLogit model. This section introduces imprecise logit-normal model according to Bickis (2009) who was the first to present it. The new imprecise model will be briefly referred to by ILnM. There will be a focus on the imprecise bivariate logit-normal model (IBLnM).

Bickis (2009) applied the imprecise logit-normal model to the estimation of hazard functions. The application of imprecise models to estimate hazard functions was also considered by Bickis and Bickis (2007), where lower and upper probabilities are derived for the imminent recurrence of pandemic influenza. Bickis and Bickis (2007) made the application using data given in Patterson (1987).

In ILnM, a family of logit-normal pdf's will be built. In Bickis (2009), the mode is considered as an estimator of the parameter θ . If the derivative of the log of the logit-normal density function is taken, then a solution for the mode can be found. The derivative of (B.6) is

$$\begin{aligned} \frac{\partial \{\log [f(\theta|\mu, \sigma)]\}}{\partial \theta} = & \frac{\partial}{\partial \theta} \left[\log \left(\frac{1}{\sigma \sqrt{2\pi}} \right) - \frac{\mu^2}{2\sigma^2} \right] + \frac{\partial}{\partial \theta} \left\{ -\frac{1}{2\sigma^2} \left[\log \left(\frac{\theta}{1-\theta} \right) \right]^2 \right\} \\ & + \frac{\partial}{\partial \theta} \left[\left(\frac{\mu}{\sigma^2} - 1 \right) \log(\theta) \right] + \frac{\partial}{\partial \theta} \left[-\left(\frac{\mu}{\sigma^2} + 1 \right) \log(1-\theta) \right] \\ = & -\frac{1}{2\sigma^2} \frac{\partial}{\partial \theta} \left\{ \left[\log \left(\frac{\theta}{1-\theta} \right) \right]^2 \right\} + \left(\frac{\mu}{\sigma^2} - 1 \right) \frac{\partial}{\partial \theta} [\log(\theta)] \\ & - \left(\frac{\mu}{\sigma^2} + 1 \right) \frac{\partial}{\partial \theta} [\log(1-\theta)] \end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{\sigma^2} \left[\log \left(\frac{\theta}{1-\theta} \right) \right] \frac{\frac{1}{(1-\theta)^2}}{\frac{1}{1-\theta}} + \left(\frac{\mu}{\sigma^2} - 1 \right) \frac{1}{\theta} + \left(\frac{\mu}{\sigma^2} + 1 \right) \frac{1}{1-\theta} \\
&= -\frac{1}{\sigma^2} \left[\log \left(\frac{\theta}{1-\theta} \right) \right] \frac{1}{\theta(1-\theta)} + \frac{\mu}{\sigma^2} \left(\frac{1}{\theta} + \frac{1}{1-\theta} \right) - \frac{1}{\theta} + \frac{1}{1-\theta} \\
&= -\frac{1}{\sigma^2} \left[\log \left(\frac{\theta}{1-\theta} \right) \right] \frac{1}{\theta(1-\theta)} + \frac{\mu}{\sigma^2} \left[\frac{1}{\theta(1-\theta)} \right] + \frac{2\theta-1}{\theta(1-\theta)} \\
&= \frac{-\frac{1}{\sigma^2} \left[\log \left(\frac{\theta}{1-\theta} \right) \right]}{\theta(1-\theta)} + \frac{\frac{\mu}{\sigma^2}}{\theta(1-\theta)} + \frac{2\theta-1}{\theta(1-\theta)} = 0 \\
&\quad -\frac{1}{\sigma^2} \left[\log \left(\frac{\theta}{1-\theta} \right) \right] + \frac{\mu}{\sigma^2} + 2\theta - 1 = 0 \\
\Rightarrow \quad &\frac{\mu}{\sigma^2} + 2\theta - 1 = \frac{1}{\sigma^2} \left[\log \left(\frac{\theta}{1-\theta} \right) \right], \tag{3.68}
\end{aligned}$$

and this results in having the mode for the logit-normal pdf in (2.40) to satisfy

$$\log \left(\frac{\theta}{1-\theta} \right) = \sigma^2 (2\theta - 1) + \mu, \tag{3.69}$$

where the line in (3.69) intersects the θ -axis when

$$0 = \sigma^2 (2\theta - 1) + \mu, \tag{3.70}$$

which means that

$$\theta_0 = \frac{1}{2} \cdot \left(\frac{-\mu}{\sigma^2} + 1 \right) = \frac{\left(-1 + \frac{\mu}{\sigma^2} \right)}{-2}. \tag{3.71}$$

Based on (B.7), (B.9) and (3.69), there is

$$\begin{aligned}
\log \left(\frac{\theta}{1-\theta} \right) &= \frac{(\eta_2^* + \eta_3^*)\theta - \eta_2^*}{2\eta_1^*} \\
&= \frac{(\eta_2 + y + \eta_3 + n - y)\theta - \eta_2 - y}{2\eta_1} \\
&= \frac{\left(\frac{\mu}{\sigma^2} - 1 + y - \frac{\mu}{\sigma^2} - 1 + n - y \right) \theta - \frac{\mu}{\sigma^2} + 1 - y}{-2\frac{1}{2\sigma^2}} \\
&= -\sigma^2 \left[\left(\frac{\mu}{\sigma^2} - 1 + y - \frac{\mu}{\sigma^2} - 1 + n - y \right) \theta - \frac{\mu}{\sigma^2} + 1 - y \right] \\
&= -\sigma^2 (n-2)\theta + \sigma^2 \left(\frac{\mu}{\sigma^2} + y - 1 \right) \\
&= -(n-2)\sigma^2\theta + \mu + (y-1)\sigma^2. \tag{3.72}
\end{aligned}$$

The line in (3.72) intersects the θ -axis at

$$\theta_0 = \frac{y-1}{n-2} + \frac{\mu}{(n-2)\sigma^2} = \frac{y-1 + \frac{\mu}{\sigma^2}}{n-2}, \quad n > 2, \quad (3.73)$$

According to Bickis (2009), the posterior modes must lie in

$$\begin{aligned} & \left\{ \inf_{\pi(\theta|\mu,\sigma) \in \mathcal{A}_{\text{prior}}} \text{mode} [\pi(\theta|y)], \sup_{\pi(\theta|\mu,\sigma) \in \mathcal{A}_{\text{prior}}} \text{mode} [\pi(\theta|y)] \right\} \\ &= \left\{ \inf_{\pi(\theta|\mu,\sigma) \in \mathcal{A}_{\text{prior}}} \left[\frac{y-1 + \frac{\mu}{\sigma^2}}{n-2} \right], \sup_{\pi(\theta|\mu,\sigma) \in \mathcal{A}_{\text{prior}}} \left[\frac{y-1 + \frac{\mu}{\sigma^2}}{n-2} \right] \right\}, \end{aligned} \quad (3.74)$$

where the class of logit-normal distributions is

$$\mathcal{A}_{\text{prior}} = \left\{ \text{logit-normal}(\mu, \sigma) : \sigma = \sigma_0 + \tau |\mu|^{\frac{1}{2}}, -\infty < \mu < \infty \right\}, \quad (3.75)$$

such that $\sigma_0 > 0$ and $\tau < 1$. A simple comparison of (3.71) to (3.73) shows that (3.73) updates (3.71).

Bickis (2009) uses (3.75) to treat the logit-normal model imprecisely. Bickis broke the habit by looking at an imprecise prior and posterior distributions other than the usual used IDM model as in Walley (1996) and Bernard (2005).

In this section, to suggest a class of priors in imprecise logit-normal prior for 2-parameter ImpLogit model, the bivariate case denoted by IBLnM is used. Under any selected pair of θ_j, θ_k , let the class of logit-normal priors be

$$\begin{aligned} \mathcal{A}_{\text{prior}} = & \\ \{ \text{logit-normal}(\mu_j, \mu_k, \sigma_j, \sigma_k, \sigma_{jk}, \sigma_{kj}) : (\mu_j, \mu_k) \in (\boldsymbol{\mu}), (\sigma_j, \sigma_k, \sigma_{jk}, \sigma_{kj}) \in (\boldsymbol{\sigma}) \}. \end{aligned} \quad (3.76)$$

Imprecision in bivariate logit-normal model is taken into account by varying the values of elements in $\boldsymbol{\mu}_{2 \times 1}$ and $\boldsymbol{\Sigma}_{2 \times 2}$. This requires allowing the location parameters μ_j and μ_k to take values over intervals that belong to the line of real numbers, \mathbb{R} . The same can be done for the scale parameters σ_j and σ_k but over intervals of positive values.

The correlation form in (3.67) is used to model the covariance matrix with $w = 1$, H_j and H_k values as in (3.57) and (3.58). Regression parameters estimates are found using the posterior of bivariate logit-normal distribution given in (2.49), under a set of prior believed values of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$.

Tables D.7, D.8 and D.9 provide the computed interval estimates for data in Table 2.1. The location parameters are fixed as $\mu_j \in \{-0.5, -1\}$, $\mu_k \in \{0.5, 1\}$. Recall that

$$E \left[\log \left(\frac{\theta_i}{1 - \theta_i} \right) \right] = \mu_i, \quad \forall i = 1, \dots, m. \quad (3.77)$$

Therefore, having $\mu_{j1} = -0.5$ and $\mu_{j2} = -1$ results in

$$\begin{aligned} E \left[\log \left(\frac{\theta_{j1}}{1 - \theta_{j1}} \right) \right] &\approx \log \left[\frac{E(\theta_{j1})}{1 - E(\theta_{j1})} \right] = -0.5, \\ \implies E(\theta_{j1}) &\approx \frac{e^{-0.5}}{1 + e^{-0.5}} \approx 0.38, \\ \text{and } E(\theta_{j2}) &\approx \frac{e^{-1}}{1 + e^{-1}} \approx 0.27. \end{aligned} \quad (3.78)$$

The same thing is for $\mu_{k1} = 0.5$ and $\mu_{k2} = 1$, where

$$\begin{aligned} E(\theta_{k1}) &\approx \frac{e^{0.5}}{1 + e^{0.5}} \approx 0.62, \\ \text{and } E(\theta_{k2}) &\approx \frac{e^1}{1 + e^1} \approx 0.73. \end{aligned} \quad (3.79)$$

Values in (3.78) and (3.79) are located on both sides of $\theta = 0.5$ (median). Tables D.7, D.8 and D.9 show that increasing the scale parameter values in IBLnM results in producing shorter interval estimates for β_0 and β_1 .

Figure D.2 shows ImpLogit plots under IBLnM. The impact of location and scale parameters on the produced imprecision under IBLnM will be paid attention in Chapter 4. Section 3.7 finalizes this chapter by fitting ImpLogit model under IDM with a small sample size.

3.7 ImpLogit with Smaller Sample

A re-look at tables 3.1, 3.2 and 3.3 for data in Table 2.1 shows that interval estimates lengths are near to each other under different values of ν and different selections of θ_j and θ_k . This can be due to having a total number of binary observations over 6 binomial dose levels as $49+50+48+46+49+50 = 292$. Therefore, this sample size will be decreased by taking a smaller sample from data given in Table 2.1. To take such a small sample, a random variable say $u \sim \text{bernoulli}(0.1, 1)$, will be used to decide whether to select each

binary observation out of 292 observations. The main goal behind searching 2-parameter ImpLogit behaviour with smaller sample is to see if imprecision of regression parameters estimates can be impacted. Table 3.13 gives the smaller sample.

Fitting 2-parameter increasing ImpLogit model under IDM with this small sample results in a longer interval estimates as being seen in tables D.10, D.11 and D.12 in comparison to tables 3.1, 3.2 and 3.3. The total sum of of binary trials in Table 3.13 is 37.

Table 3.13: Smaller size binomial data sample taken from the population data in Table 2.1. Dose is given in mg/l, n is the number of experimented insects and y_{obs} is the number of affected insects.

	Dose	n	y_{obs}
1	0	5	0
2	2.6	6	1
3	3.8	7	2
4	5.1	6	3
5	7.7	6	4
6	10.2	7	7

CHAPTER 4

SIMULATION STUDY

4.1 Motivation for the Simulation Study

A simulation study is necessary to have a deeper knowledge of 2-parameter ImpLogit model. Several points of interest can arise and require more investigation. For example, the sample size is an important factor that relates to the imprecision of β_0 and β_1 estimates. The number of binomial observations (m) is a point that deserves to be paid attention.

Furthermore, the locations of the covariate x_i distinct values can play a strong role in increasing or decreasing the imprecision of parameters of interest. This comes by searching how to shorten the interval estimates of regression parameters. A simulation study can search the conditions under which the true values of regression parameters do belong to the produced interval estimates. The location and distance among the selected values of the covariate x_i have to be considered. It is important to know how far the selected values of x_i should be from each other in order to control the lengths of interval estimates.

The proposed simulation study in this chapter studies differences in ImpLogit fit and performance under both of IDM and IBLnM imprecise priors. It is interesting to study the effect of changing the imprecise priors parameters on the produced imprecision of regression parameters. Also, the relation between the values of parameters in imprecise priors and lengths of the produced interval estimates is important to be figured out.

To start the simulation study, let $\beta_0 = 0$ and $\beta_1 = 0.4$. Having true $\beta_0 = 0$ makes the true median of the logistic distribution to be

$$\log\left(\frac{0.5}{1-0.5}\right) = \beta_0 + x_i(\beta_1) = (0.4)x_i \implies x_{\text{median}} = 0.$$

Also, having $\beta_1 = 0.4$ makes the true θ at $x_i = -1$ and $x_i = 1$ as

$$\left. \begin{aligned} \theta_i = \theta(x_i = -1) &= \frac{e^{0+0.4(-1)}}{1 + e^{0+0.4(-1)}} = \frac{e^{-0.4}}{1 + e^{-0.4}} \approx 0.4, \\ \text{and } \theta_i = \theta(x_i = 1) &= \frac{e^{0+0.4(1)}}{1 + e^{0+0.4(1)}} = \frac{e^{0.4}}{1 + e^{0.4}} \approx 0.6. \end{aligned} \right\} \quad (4.1)$$

So, the true θ changes within 0.4 to 0.6 over covariate range of -1 to 1, where the true median value is at $x_i = 0$. In other words, the x_i 's can be scaled such that $x_i = 0$ corresponds to a response of 0.5 and $x_i = -1$ and 1 correspond to responses of 0.4 and 0.6, respectively.

For the simulated data, each binomial observation y_i comes from

$$y_i \sim \text{bin} \left(n_i, \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right), \quad \beta_0 = 0, \quad \beta_1 = 0.4, \quad i = 1, \dots, m, \quad (4.2)$$

$$n_1 = n_2 = \dots = n_m, \quad (4.3)$$

where the design is balanced and the total number of trials over all binomial observations is fixed as $\sum_{i=1}^m n_i = 40$. Three combinations of n_i and m are chosen, $n_i = 20$ with $m = 2$, $n_i = 4$ with $m = 10$ and $n_i = 2$ with $m = 20$, such that $n_i \times m = 40, \forall i$.

Since having true $\beta_1 = 0.4$ refers to an increasing logistic function, then the induced prior distribution under IDM is modelled as in (2.11) in Section 2.1. This also requires to use the correlation structure given in (2.23) in the induced prior distribution under IBLnM.

The following sections will present simulations under different designs. Sections 4.2, 4.3 and 4.4 describe simulating ImpLogit model with two covariate fixed values, say x_1 and x_2 , on different locations from $x = 0$. Having only two fixed covariate values leads to have only one selected pair of θ_j and θ_k at x_1 and x_2 , respectively. Both of x_1 and x_2 will be selected to surround or to be on the right and left sides of the true median value. Fixing only two values allows to search the relation between the selected distance between x_1 and x_2 values and the produced interval estimates. Sections 4.5, 4.6 and 4.7 are for ImpLogit simulation with multiple covariate fixed values, x_1, \dots, x_m .

For each design, ImpLogit model will be fitted over 2000 generated samples. Having 2000 generated samples will result in computing 2000 single interval estimates for β_0 and β_1 . Averages of 2000 lower estimates and 2000 upper estimates will be calculated to find

what will be called as the averaged interval estimate. From here and on, an averaged interval estimate is abbreviated by AIE.

The simulation study results will be plotted graphically from page 108 to page 119 for the case of two fixed covariate values and pages 120 and 129 for the case of multiple fixed values.

4.2 Symmetric Two Values

In sections 4.2, 4.3 and 4.4, the covariate x_i will be fixed at the following cases,

$$\left. \begin{array}{l} x_1 = -0.5 \text{ and } x_2 = 0.5, \\ x_1 = -1 \text{ and } x_2 = 1, \\ x_1 = -2.5 \text{ and } x_2 = 2.5. \end{array} \right\} \quad (4.4)$$

Start with $x_1 = -0.5$ and $x_2 = 0.5$ so that they are located on similar distance from $x = 0$. Table 4.2 (page 104) gives the simulation computations obtained for lower and upper estimates of regression parameters.

Distance between x_1 and x_2 can be lengthened in order to investigate its impact on AIEs. Tables 4.3 and 4.4 give computational results when $x_1 = -1$, $x_2 = 1$ and $x_1 = -2.5$, $x_2 = 2.5$, respectively. The AIEs in tables 4.2, 4.3 and 4.4 are plotted in figures 4.1 and 4.2 for β_0 and β_1 , respectively. Lower and upper estimates are solidly marked by a closed circle \bullet when $x_1 = -0.5$ and $x_2 = 0.5$, a closed square \blacksquare when $x_1 = -1$ and $x_2 = 1$ and a closed triangle \blacktriangle when $x_1 = -2.5$ and $x_2 = 2.5$. The true values of β_0 and β_1 are always marked by “ \times ”.

Figures 4.1 and 4.2 provide useful information about the behaviour of AIEs of β_0 and β_1 . Both of H_j and H_k are assigned values that were used in (3.57) and (3.58). The top plots in both figures are under IDM where $\nu = 1$ and $\nu = 2$, respectively. This aims to navigate the impact of changing the value of ν on the AIEs lengths of regression parameters.

The top-left plots of both figures, where $\nu = 1$, show that true values of β_0 and β_1 are falling inside AIEs for all choices of x_i . A similar thing happens in the top-right plot when $\nu = 2$. Increasing ν does not necessarily result in a noticeable difference in the AIEs

of β_0 in terms of length and values of lower and upper estimates. A bigger difference can be seen for β_1 in Figure 4.2. AIEs of β_1 in the top-right plot are looking longer than the corresponding ones in the top-left plot. One more important thing that a shorter distance between x_1 and x_2 produces longer interval estimates which can clearly be noticed in top plots of both figures. This indicates, tentatively, that a shorter distance between x_1 and x_2 results in more imprecision.

For each single (not averaged) interval estimate, the lower value can be plotted versus the upper value. Figure 4.3 gives 6 different plots in which the left column corresponds to the top-left plot of Figure 4.1 while the right column is for the top-right plot. Each plot in Figure 4.3 includes 2000 points. Each point locates above the dashed diagonal line because the lower estimate must be less than the upper estimate in any single interval estimate. Left and right columns of Figure 4.4 give plots that correspond to the AIEs in top-left and top-right plots of Figure 4.2, respectively.

In each plot of figures 4.3 and 4.4, the diagonal, horizontal and vertical lines divide the dotted area into three smaller areas. The smaller areas look like a square and two equal triangles. The proportion of points out of all points in the smaller square in each plot is given as percentage value in top plots of figures 4.1 and 4.2. Any point in that square represents a single interval estimate that includes the true regression parameter value.

Percentage values in top plots are equal or higher than 95% and 90% for estimates of β_0 and β_1 , respectively, which matches with plots in figures 4.3 and 4.4 where few points are located in the top-right and bottom-left triangles. In other words, few cases of AIEs with extreme near high or low values of lower and upper estimates were found because most points are in the top-left square. This means that true values of β_0 and β_1 were included in greater than or equal to 95% and 90% out of all AIEs of β_0 and β_1 , respectively, under IDM. This can be noticed in top plots of figures 4.1 and 4.2 where “ \times ” is almost at the middle of all AIEs. Percentage values will be presented in all figures of this simulation study.

In middle-left plots of figures 4.1 and 4.2, the fixed prior values of location and scale

parameters in IBLnM are

$$\mu_j \in \{-0.35, -0.45\} , \mu_k \in \{0.35, 0.45\} , \quad (4.5)$$

$$\text{and } \sigma_j = \sigma_k \in \{0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\} , \quad (4.6)$$

while the middle-right plot shows AIEs when

$$\mu_j \in \{-0.5, -1\} , \mu_k \in \{0.5, 1\} , \quad (4.7)$$

with similar values of σ_j and σ_k as in (4.6). Having the location parameters values as in (4.5) and (4.7) means that

$$\left. \begin{aligned} \text{if } \mu_i = -0.35 &\implies E(\theta_i) \approx \frac{e^{-0.35}}{1 + e^{-0.35}} \approx 0.41, \\ \text{if } \mu_i = -0.45 &\implies E(\theta_i) \approx \frac{e^{-0.45}}{1 + e^{-0.45}} \approx 0.39, \\ \text{if } \mu_i = 0.35 &\implies E(\theta_i) \approx \frac{e^{0.35}}{1 + e^{0.35}} \approx 0.59, \\ \text{if } \mu_i = 0.45 &\implies E(\theta_i) \approx \frac{e^{0.45}}{1 + e^{0.45}} \approx 0.61, \end{aligned} \right\} \quad (4.8)$$

which means that, in parallel with (3.78), (3.79) and (4.8), the prior approximate belief about θ ranges between 0.27 and 0.73. This prior belief of θ range can be compared to the case of having $x_1 = -2.5$ and $x_2 = 2.5$ from (4.4), provided that $\beta_0 = 0$, $\beta_1 = 0.4$, as

$$\left. \begin{aligned} \theta_i = \theta(x_i = -2.5) &= \frac{e^{0+0.4(-2.5)}}{1 + e^{0+0.4(-2.5)}} = \frac{e^{-1}}{1 + e^{-1}} \approx 0.27, \\ \text{and } \theta_i = \theta(x_i = 2.5) &= \frac{e^{0+0.4(2.5)}}{1 + e^{0+0.4(2.5)}} = \frac{e}{1 + e} \approx 0.73, \end{aligned} \right\} \quad (4.9)$$

which shows the closeness of both ranges.

Middle plots of figures 4.1 and 4.2 show that all AIEs included the true values of β_0 and β_1 , with less noticeable difference in lengths of AIEs of β_0 . The imprecise model behaviour, in terms of AIEs lengths, changes in the middle plots of Figure 4.2 for β_1 . In Figure 4.2, imprecision of β_1 is impacted and reduced under longer distance between x_1 and x_2 which is the same in top plots. The impact of covariate values range on lengths of AIEs is less on β_0 in comparison to β_1 .

The bottom plots in both figures are for IBLnM with values of μ_j and μ_k similar to those in the middle plots but with larger values of σ_j and σ_k . The scaling prior parameters

are fixed as

$$\sigma_j = \sigma_k \in \{1.0, 1.08, 1.16, 1.24, 1.32, 1.40, 1.48\}. \quad (4.10)$$

In bottom plots, larger σ_j and σ_k lead to shorter AIEs in β_0 and β_1 , but with stronger impact on β_1 . The scale parameters σ_j and σ_k in IBLnM tend to play a strong role as for ν in IDM. The true values of both of β_0 and β_1 are always included in the AIEs which is quite visibly seen in figures 4.1 and 4.2. The relation between lengths of AIEs and the distance between x_1 and x_2 agrees with the tentative results found in the top and middle plots in both figures. Generally speaking, a shorter distance between x_1 and x_2 does not necessarily produce less imprecision.

Increasing the difference between μ_j and μ_k results in changing the lengths of the AIEs. Figures 4.1 and 4.2 acknowledge that increasing σ_j and σ_k shortens the AIEs, while changing μ_j farther to the left and μ_k to the right, lengthens the AIEs.

This section summarizes as the following : imprecision of regression parameters estimates in ImpLogit model is increased if the distance between x_1 and x_2 is shortened. If IDM is applied, a larger ν results in longer AIEs. If IBLnM is considered as an imprecise prior, then smaller σ_j and σ_k lead to have longer AIEs. Dragging μ_j and μ_k farther from each other makes longer AIEs. If x_i 's is symmetrically allocated around the true median value, then true values of β_0 and β_1 are always included in the AIEs under both imprecise priors.

It is important to note there is no change in the values of ν , μ_i and σ_i among all figures in the rest of this simulation study. Section 4.3 suggests to drag both covariate fixed values to the right and left sides of $x = 0$.

4.3 Right and Left-Shifted Two Values

The symmetric design is not always the case where x_1 and x_2 can be allocated to the right or left of the true median $x = 0$. Figures 4.5 and 4.6 present simulated AIEs when $x_1 = 1$, $x_2 = 2$, 3 and 4. Right-shifted x_1 and x_2 design gives AIEs for β_0 that include its true value as being seen in Figure 4.5. The only thing that can be noticed is the change of “ \times ” location on the AIEs plots, where “ \times ” tends to move to the right side of each AIE.

Figure 4.6 shows that β_1 true value is always included in the AIEs. The imprecision produced in case of $x_1 = 1$ and $x_2 = 2$ is larger than the case of $x_1 = 1$ and $x_2 = 3$ or 4. That is, a shorter range results in larger imprecision even for right-shifted covariate values. This does not conflict with Section 4.2 regarding the relation between the lengths of the AIEs and the distance between x_1 and x_2 . AIEs of β_1 are sometimes longer compared to those found in the symmetric design. Therefore, shifting x_1 and x_2 may create longer AIEs of β_1 which stands against the main goal of fitting and using ImpLogit model.

Increasing ν in the top-right plot of Figure 4.6 under IDM makes longer AIEs for β_1 . It can be seen by looking at top plots in both figures that increasing ν has stronger impact on β_1 than β_0 . Despite that imprecision is less when $x_1 = 1$ and $x_2 = 3$ or 4, the AIEs included β_0 and β_1 true values in all plots.

Comparing the middle and bottom plots of Figure 4.6 shows that larger values of σ_j and σ_k result in shorter AIEs for β_1 . Enlarging σ_j and σ_k may and may not decrease imprecision of β_0 in Figure 4.5 which is different from the previous design where AIEs lengths in β_0 and β_1 were decreased with larger σ_j and σ_k . Shifting the location parameters farther to both sides of $\mu = 0$ lengthens AIEs which matches with the case of symmetric located x_1 and x_2 . Again, scale parameters in IBLnM prior are playing a strong role in lengthening or shortening the AIEs for β_1 .

ImpLogit model with left-shifted covariate values is taken into account. The covariate selections are $x_1 = -2, -3, -4$ and $x_2 = -1$. Plots of AIEs are given in figures 4.7 and 4.8. In both figures, ImpLogit model behaviour is quite similar to the right-shifted case. Left-shifted design matches with both of symmetric and right-shifted ones in letting a wider distance between x_1 and x_2 to reduce the produced imprecision. True values of both of β_1 and β_0 are always included in the AIEs. The true value of β_0 marked by “ \times ” tends to move toward the left side of each AIE plot in Figure 4.7 (it is to the right side in Figure 4.5).

4.4 Extremely Shifted Two Values

Both of x_1 and x_2 are now suggested to be located farther on both sides of $x = 0$. For the right side, there is $x_1 = 3.5, x_2 = 4.5$ and $x_1 = 3, x_2 = 5$ where the AIEs plots are

given in figures 4.9 and 4.10. This new situation finds that true β_0 fails to belong to any of the computed AIEs under both of IDM and IBLnM as being figured out in Figure 4.9. If $x_1 = 3.5$ and $x_2 = 4.5$, then longer AIEs are produced in comparison to those when $x_1 = 3$ and $x_2 = 5$. This happens in all plots of figures 4.9 and 4.10. This gives a stronger evidence that, with two fixed covariate values, longer AIEs come with shorter fixed ranges. In Figure 4.10, β_1 true value is included in the AIEs only if $x_1 = 3.5$ and $x_2 = 4.5$.

The last attempt with two fixed covariate values is by considering $x_1 = -4.5$, $x_2 = -3.5$ and $x_1 = -5$, $x_2 = -3$. Having $x_1 = -4.5$ and $x_2 = -3.5$ produces longer AIEs as being seen in figures 4.11 and 4.12 than if $x_1 = -5$, $x_2 = -3$. True β_0 is out of all AIEs in Figure 4.11. This is similar to the case of extreme right-shifted values. AIEs of β_1 do include the true value only when $x_1 = -4.5$, $x_2 = -3.5$ but not if $x_1 = -5$, $x_2 = -3$ which can be seen in Figure 4.12. Generally, with farther locations of x_1 and x_2 (not around $x = 0$ or near to $x = 0$ from both sides), true values of both of β_0 and β_1 tend to escape from the corresponding AIEs.

The impact of ν over the current covariate locations does not change from that in the previous two sections. Larger ν gives longer AIEs under IDM. Less values of scale parameters and smaller μ_j with larger μ_k under IBLnM give longer AIEs.

The simulation study can be extended and enhanced by fixing more than two values for x_i . Section 4.5 starts by fixing multiple values around $x = 0$ symmetrically.

4.5 Symmetric Covariate Multiple Values

To simulate ImpLogit under multiple values of the covariate x_i , the following fixed values are kept : $\beta_0 = 0$ and $\beta_1 = 0.4$, $m \times n_i = 40$, $n = n_i$, $\forall i = 1, \dots, m$, but $m \in \{5, 10, 20\}$ instead of $m \in \{2, 10, 20\}$. The covariate ranges are $x_i \in [-0.5, 0.5]$, $x_i \in [-1, 1]$ and $x_i \in [-2.5, 2.5]$. Table 4.5 presents the multiple covariate values. They are selected to surround $x = 0$ in order to have symmetrical circulation around the true median value.

If $m = 5$, the selected binomial parameters that induce the regression parameters prior distribution are θ_2 and θ_4 (corresponding to x_2 and x_4). The selected parameters become θ_3 and θ_8 when $m = 10$ and θ_5 and θ_{16} if $m = 20$. The given selections do not change in sections 4.6 and 4.7.

The AIEs are plotted in figures 4.13 and 4.14. Having multiple covariate values finds that short ranges of x_i produce longer AIEs. The true parameters values are always falling within the AIEs for all plots and all values of m . Larger ν and smaller σ_j and σ_k result in longer AIEs. Simulation results under multiple covariate values don't show a difference from Section 4.2. The main new thing is that AIEs are longer when $m = 10$ or 20.

4.6 Right and Left-Shifted Multiple Values

Figures 4.15 and 4.16 give plots when $x_i \in [1, 2]$, $x_i \in [1, 3]$ and $x_i \in [1, 4]$. AIEs always include true β_0 and β_1 . Figures 4.15 and 4.16 shows that enlarging m results in longer AIEs. Note that β_0 true value is located to the right side of each AIE in Figure 4.15 which has happened in Figure 4.5. Longer AIEs are obtained under $x_i \in [1, 2]$ for both of β_0 and β_1 .

Figures 4.17 and 4.18 are for left-shifted covariate multiple values and the selected covariate ranges are $x_i \in [-4, -1]$, $x_i \in [-3, -1]$ and $x_i \in [-2, -1]$. Having $x_i \in [-2, -1]$ produced longer AIEs for both of β_0 and β_1 than the other two cases of covariate values ranges. True values of regression parameters are always included in AIEs. Larger m gives longer AIEs.

4.7 Extremely Shifted Multiple Values

Figures 4.19 and 4.20 are with $x_i \in [3.5, 4.5]$ and $x_i \in [3, 5]$. In both figures, true β_0 falls outside all AIEs. Larger m gives longer AIEs for both of β_0 and β_1 . The true β_1 is out of AIEs when $x_i \in [3, 5]$.

Figures 4.21 and 4.22 provide plots when $x_i \in [-4.5, -3.5]$ and $x_i \in [-5, -3]$. Having $x_i \in [-4.5, -3.5]$ results in longer AIEs than if $x_i \in [-5, -3]$. It is important to see that “ \times ” is out of AIEs of β_0 in all plots. Figure 4.22 shows that true β_1 is not included when $x_i \in [-5, -3]$.

4.8 Closer θ_j and θ_k

The relation between the closeness of the sequential orders of the selected pair θ_j and θ_k and the imprecision of $\hat{\beta}_0$ and $\hat{\beta}_1$ can be investigated. For example, if $m = 20$, $n_i = 2$, $\forall i$, the question is : does ImpLogit fit give different results when the selected pair is $\theta_5 = \theta(x_5)$ and $\theta_{16} = \theta(x_{16})$ than if it is $\theta_{10} = \theta(x_{10})$ and $\theta_{11} = \theta(x_{11})$? In other words : can the selected priors retard the role of x_i 's values range on including or firing the true regression parameters values from their AIEs?

To answer the last question, figures E.1 and E.2 (in Appendix E) present plots of AIEs under the symmetric design when $m = 20$ and $n_i = 2$, $\forall i$ and different selected θ_j and θ_k . The x_i 's values are similar to those in Table 4.5 when $m = 20$. That is, Figure E.1 corresponds to Figure 4.13, while Figure E.2 corresponds to Figure 4.14 with similar corresponding values of ν in IDM and μ_j , μ_k , σ_j and σ_k in IBLnM. The difference in figures is simplified as : Figure 4.13 shows AIEs taking into account different values of m , while Figure E.1 shows AIEs with fixed $m = 20$ and selections of θ_5 and θ_{16} , θ_8 and θ_{13} and finally θ_{10} and θ_{11} . Therefore, both figures are having identical AIEs only when $m = 20$ and θ_5 and θ_{16} (given in black in the top three AIEs of each plot). The same situation can be seen in figures E.2 and 4.14. It is important to note that graphs of this section and sections 4.9 and 4.10 are given in Appendix E.

Selecting different pairs of θ_j and θ_k does not show a clear difference in AIEs lengths under the symmetric design. The location of the true parameters values kept being inside

all computed AIEs, but the proportion of single interval estimates that include true values of β_0 and β_1 is increased compared to percentage values in figures 4.13 and 4.14.

The same thing is applied for extreme x_i 's where figures E.3 and E.4 correspond to figures 4.19 and 4.20, respectively. In Figure E.3, selecting θ_8 , θ_{13} or θ_{10} , θ_{11} does not result in true β_0 to fall in any of the computed AIEs. The main thing is that the proportion of single interval estimates that include true β_0 is increased. AIEs with θ_5 and θ_{16} are similar to those in Figure 4.19. Selecting θ_8 , θ_{13} or θ_{10} , θ_{11} makes true β_1 nearer to the AIEs but not brought inside any AIE when $x_i \in [3, 5]$.

4.9 Optimal Frequentist Design

The study of x_i 's values design in ImpLogit model can be developed using the Fisher information matrix (FIM). This enables to find the x_i 's at which the asymptotic variances of $\hat{\beta}_0$ and $\hat{\beta}_1$ are taken to their minimum values. The main goal is to build a design from FIM and compare it to the previous ImpLogit designs. The FIM of β_0 and β_1 is given as

$$\begin{bmatrix} \sum_{i=1}^m n_i \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right) & \sum_{i=1}^m x_i n_i \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right) \\ \sum_{i=1}^m x_i n_i \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right) & \sum_{i=1}^m n_i x_i^2 \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right) \end{bmatrix}. \quad (4.11)$$

The matrix in (4.11) is given directly, with no derivation, in Casella and Berger (2002). Appendix F shows the derivation of (4.11). Now, (4.11) is used with the true regression parameters values that were used in previous sections, $\beta_0 = 0$ and $\beta_1 = 0.4$, to become

$$\begin{bmatrix} \sum_{i=1}^m n_i \frac{e^{0.4x_i}}{1 + e^{0.4x_i}} \left(\frac{1}{1 + e^{0.4x_i}} \right) & \sum_{i=1}^m x_i n_i \frac{e^{0.4x_i}}{1 + e^{0.4x_i}} \left(\frac{1}{1 + e^{0.4x_i}} \right) \\ \sum_{i=1}^m x_i n_i \frac{e^{0.4x_i}}{1 + e^{0.4x_i}} \left(\frac{1}{1 + e^{0.4x_i}} \right) & \sum_{i=1}^m n_i x_i^2 \frac{e^{0.4x_i}}{1 + e^{0.4x_i}} \left(\frac{1}{1 + e^{0.4x_i}} \right) \end{bmatrix}. \quad (4.12)$$

The inverse of (4.12) has to be found in order to find approximate estimates of the asymptotic variances of $\hat{\beta}_0$ and $\hat{\beta}_1$. The determinant of (4.12) is

$$D(\mathbf{x}) = \left[\sum_{i=1}^m n_i \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2} \right] \left[\sum_{i=1}^m n_i x_i^2 \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2} \right] - \left[\sum_{i=1}^m x_i n_i \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2} \right]^2, \quad (4.13)$$

\Rightarrow

$$I^{-1} = \begin{bmatrix} \frac{\sum_{i=1}^m n_i x_i^2 \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2}}{D(x_1, \dots, x_m)} & -\frac{\sum_{i=1}^m x_i n_i \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2}}{D(x_1, \dots, x_m)} \\ -\frac{\sum_{i=1}^m x_i n_i \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2}}{D(x_1, \dots, x_m)} & \frac{\sum_{i=1}^m n_i \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2}}{D(x_1, \dots, x_m)} \end{bmatrix}. \quad (4.14)$$

Now, x_i 's in (4.14) are considered as variables, then derivatives with respect to x_i are taken and equated to zero value to be solved numerically as

$$\frac{\partial}{\partial x_i} \left[\frac{\sum_{i=1}^m n_i x_i^2 \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2}}{D(x_1, \dots, x_m)} \right] = 0, \quad (4.15)$$

and

$$\frac{\partial}{\partial x_i} \left[\frac{\sum_{i=1}^m n_i \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2}}{D(x_1, \dots, x_m)} \right] = 0. \quad (4.16)$$

Maximizing the determinant function in (4.13) or solving any of equations (4.15) and (4.16) can give x_i 's values that optimize the estimates of the predictive probabilities θ_i , $\forall i$. The covariate values that come from maximizing (4.13) or solving (4.15) or (4.16) are not necessarily the same. The values can be compared to x_i 's ranges in ImpLogit design. In other way, the main question is : do such x_i 's values (provided that $\beta_0 = 0$ and $\beta_1 = 0.4$) match with those that reduce imprecision in ImpLogit model according to previous ImpLogit designs?

To answer the last question : Table 4.1 presents the computed values of $D(\mathbf{x})$ in (4.13) under different x_i 's values. Larger values of the determinant function come when

$x_1 = -2.5$, $x_2 = 2.5$ or $x_i \in [-2.5, 2.5]$ while values of $D(\mathbf{x})$ when $x_1 = -0.5$, $x_2 = 0.5$ or $x_i \in [-0.5, 0.5]$ were the smallest. This matches with previous ImpLogit designs results, because a covariate range with larger $D(\mathbf{x})$ in Table 4.1 corresponds to a covariate range in ImpLogit model with less imprecision in regression parameters estimates. Table 4.1 shows that increasing m decreases $D(\mathbf{x})$ value which means that having larger m produces longer interval estimates of β_0 and β_1 . This is not different from what was found in the previous sections of this chapter.

Table 4.1: Computed $D(\mathbf{x})$ values using (4.13) under the symmetric design provided that $\sum_{i=1}^m n_i = 40$.

Two x_i 's			Multiple x_i 's			
$m = 2$			$m = 5$	$m = 10$	$m = 20$	
$x_1 = -0.5$, $x_2 = 0.5$	24.5	$x_i \in [-0.5, 0.5]$	12.3	10.3	8.6	
$x_1 = -1$, $x_2 = 1$	92.4	$x_i \in [-1, 1]$	47.4	42	36.9	
$x_1 = -2.5$, $x_2 = 2.5$	386.6	$x_i \in [-2.5, 2.5]$	211.3	209.3	183.7	

The maximum value of the determinant function in (4.13) occurs when $x_1 = -3.85$ and $x_2 = 3.85$. This covariate range includes the range of $x_1 = -2.5$, $x_2 = 2.5$ (that produced the shortest interval estimates). Therefore, it is interesting to fit ImpLogit model when $x_1 = -3.85$ and $x_2 = 3.85$ to see the reduction of imprecision in regression parameters estimates. Figures E.5 and E.6 add to Figures 4.1 and 4.2 plots of a fourth interval estimate under the range of $x_1 = -3.85$ and $x_2 = 3.85$. The ends of the new interval estimate plots are marked by ∇ . Other plots of interval estimates (marked by \bullet , \blacksquare or \blacktriangle) are copied from Figures 4.1 and 4.2. Figures E.5 and E.6 show that under $x_1 = -3.85$ and $x_2 = 3.85$, the interval estimates are shorter for both of β_0 and β_1 but with a stronger impact on β_1 .

For $m = 2$, solving (4.15) gives x_i 's values at which the approximate asymptotic variance of $\hat{\beta}_0$ is minimized, in the inverse of FIM, $\forall i$, to be at $x_1 = x_2 = 0$. The approximate asymptotic variance of $\hat{\beta}_1$ takes its minimum value at $x_1 = -5.998$ and $x_2 = 5.998$ after solving (4.16). This covariate range includes $x_1 = -2.5$, $x_2 = 2.5$.

4.10 Multiple Parameter ImpLogit Model

ImpLogit model can be extended and developed by considering multiple parameters, $\beta_0, \beta_1, \dots, \beta_p$. $(P+1)$ -parameter ImpLogit model gives the opportunity to study the effects of p covariates on the produced imprecision of $p + 1$ regression parameters. If

$$\theta_i = \theta(\mathbf{x}_i) = \frac{e^{\mathbf{x}'_i \boldsymbol{\beta}}}{1 + e^{\mathbf{x}'_i \boldsymbol{\beta}}} = \frac{e^{\beta_0 + x_{i1}\beta_1 + \dots + x_{ip}\beta_p}}{1 + e^{\beta_0 + x_{i1}\beta_1 + \dots + x_{ip}\beta_p}}, \quad (4.17)$$

then the prior distribution of $\boldsymbol{\theta} = [\theta_j, \theta_k, \dots, \theta_t]'$ is defined as

$$f(\theta_j, \theta_k, \dots, \theta_t) = w f_1(\theta_j, \theta_k, \dots, \theta_t) + (1 - w) f_2(\theta_j, \theta_k, \dots, \theta_t), \quad w \in [0, 1]. \quad (4.18)$$

To induce a prior distribution for $\boldsymbol{\beta} = [\beta_0, \dots, \beta_p]'$, the Jacobian determinant $\left| \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\beta}} \right|$ is

$$\begin{vmatrix} \frac{\partial \theta_j}{\partial \beta_0} & \frac{\partial \theta_j}{\partial \beta_1} & \dots & \frac{\partial \theta_j}{\partial \beta_p} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial \theta_t}{\partial \beta_0} & \frac{\partial \theta_t}{\partial \beta_1} & \dots & \frac{\partial \theta_t}{\partial \beta_p} \end{vmatrix} = \begin{vmatrix} \frac{e^{\mathbf{x}'_j \boldsymbol{\beta}}}{(1 + e^{\mathbf{x}'_j \boldsymbol{\beta}})^2} & \frac{x_{j1} e^{\mathbf{x}'_j \boldsymbol{\beta}}}{(1 + e^{\mathbf{x}'_j \boldsymbol{\beta}})^2} & \dots & \frac{x_{jp} e^{\mathbf{x}'_j \boldsymbol{\beta}}}{(1 + e^{\mathbf{x}'_j \boldsymbol{\beta}})^2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \frac{e^{\mathbf{x}'_t \boldsymbol{\beta}}}{(1 + e^{\mathbf{x}'_t \boldsymbol{\beta}})^2} & \frac{x_{t1} e^{\mathbf{x}'_t \boldsymbol{\beta}}}{(1 + e^{\mathbf{x}'_t \boldsymbol{\beta}})^2} & \dots & \frac{x_{tp} e^{\mathbf{x}'_t \boldsymbol{\beta}}}{(1 + e^{\mathbf{x}'_t \boldsymbol{\beta}})^2} \end{vmatrix}. \quad (4.19)$$

For example, take the case of 3 regression parameters, β_0, β_1 and β_2 . There is

$$\begin{aligned} \theta_i &= \frac{e^{\beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2}}{1 + e^{\beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2}}, \quad -\infty < \beta_j < \infty, \quad j = 0, 1, 2, \\ -\infty < x_{ij} < \infty, \quad i &= 1, \dots, m, \quad j = 1, 2. \end{aligned} \quad (4.20)$$

Using the aggregation property of the Dirichlet distribution gives the joint prior distribution for $f_1(\theta_j, \theta_k, \theta_l)$ as

$$(\theta_j, \theta_k - \theta_j, \theta_l - \theta_k, 1 - \theta_l) \sim \text{Dir}(\nu H_j, \nu H_k - \nu H_j, \nu H_l - \nu H_k, \nu - \nu H_l), \quad (4.21)$$

and for $f_2(\theta_j, \theta_k, \theta_l)$ as

$$(1 - \theta_j, \theta_k - \theta_j, \theta_l - \theta_k, \theta_l) \sim \text{Dir}(\nu - \nu H_j^*, \nu H_j^* - \nu H_k^*, \nu H_k^* - \nu H_l^*, \nu H_l^*). \quad (4.22)$$

For $f_1(\cdot)$, let $\varphi_{jkl1} = H_j$, $\varphi_{jkl2} = H_k - H_j$, $\varphi_{jkl3} = H_l - H_k$ and $\varphi_{jkl4} = 1 - H_l$. Also, for $f_2(\cdot)$, let $\varphi_{jkl1}^* = 1 - H_j^*$, $\varphi_{jkl2}^* = H_j^* - H_k^*$, $\varphi_{jkl3}^* = H_k^* - H_l^*$ and $\varphi_{jkl4}^* = H_l^*$. If $x_{i1} = x_{i2}$, $\forall i = 1, \dots, m$, or $x_{i1} \leq x_{(i+1)1}$ and $x_{i2} \leq x_{(i+1)2}$, $\forall i = 1, \dots, m-1$, then $\theta_j, \theta_k, \theta_l$ are selected based on the sequential order. The prior distribution of $\boldsymbol{\theta} = [\theta_j, \theta_k, \theta_l]'$ is

$$f(\theta_j, \theta_k, \theta_l) = w f_1(\theta_j, \theta_k, \theta_l) + (1 - w) f_2(\theta_j, \theta_k, \theta_l), \quad w \in [0, 1], \quad (4.23)$$

where f_1 and f_2 are densities for the increasing and decreasing logistic functions. For $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2]'$ and selected $\boldsymbol{\theta} = [\theta_j, \theta_k, \theta_l]'$, to find

$$\pi(\boldsymbol{\beta}) = f[\boldsymbol{\theta}(\boldsymbol{\beta})] \left| \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\beta}} \right|, \quad (4.24)$$

the jacobian determinant is

$$\begin{vmatrix} \frac{e^{\beta_0 + x_{j1}\beta_1 + x_{j2}\beta_2}}{(1 + e^{\beta_0 + x_{j1}\beta_1 + x_{j2}\beta_2})^2} & \frac{x_{j1}e^{\beta_0 + x_{j1}\beta_1 + x_{j2}\beta_2}}{(1 + e^{\beta_0 + x_{j1}\beta_1 + x_{j2}\beta_2})^2} & \frac{x_{j2}e^{\beta_0 + x_{j1}\beta_1 + x_{j2}\beta_2}}{(1 + e^{\beta_0 + x_{j1}\beta_1 + x_{j2}\beta_2})^2} \\ \frac{e^{\beta_0 + x_{k1}\beta_1 + x_{k2}\beta_2}}{(1 + e^{\beta_0 + x_{k1}\beta_1 + x_{k2}\beta_2})^2} & \frac{x_{k1}e^{\beta_0 + x_{k1}\beta_1 + x_{k2}\beta_2}}{(1 + e^{\beta_0 + x_{k1}\beta_1 + x_{k2}\beta_2})^2} & \frac{x_{k2}e^{\beta_0 + x_{k1}\beta_1 + x_{k2}\beta_2}}{(1 + e^{\beta_0 + x_{k1}\beta_1 + x_{k2}\beta_2})^2} \\ \frac{e^{\beta_0 + x_{l1}\beta_1 + x_{l2}\beta_2}}{(1 + e^{\beta_0 + x_{l1}\beta_1 + x_{l2}\beta_2})^2} & \frac{x_{l1}e^{\beta_0 + x_{l1}\beta_1 + x_{l2}\beta_2}}{(1 + e^{\beta_0 + x_{l1}\beta_1 + x_{l2}\beta_2})^2} & \frac{x_{l2}e^{\beta_0 + x_{l1}\beta_1 + x_{l2}\beta_2}}{(1 + e^{\beta_0 + x_{l1}\beta_1 + x_{l2}\beta_2})^2} \end{vmatrix}.$$

Now, the 3-parameter ImpLogit model is simulated under IDM with specific ranges of two covariates x_{i1} and x_{i2} . The true parameters values are $\beta_0 = 0$, $\beta_1 = 0.4$ and $\beta_2 = 0.4$.

Figures E.7 and E.8 give plots of AIEs for β_0 , β_1 and β_2 . Figure E.7 gives AIEs with symmetric and similar ranges of x_{1i} and x_{2i} . If $m = 5$, then the prior distribution of regression parameters is induced after selecting θ_1 , θ_3 and θ_5 . If $m = 10$, then the selection is for θ_1 , θ_5 and θ_9 , while if $m = 20$, the selected binomial parameters are θ_1 , θ_{10} and θ_{19} . True values of all regression parameters are always included in the AIEs. Longer covariates ranges result in shortening the interval estimates of all regression parameters. Increasing m works against lengthening ranges of x_{i1} and x_{i2} by giving longer AIEs.

Figure E.8 gives plots with different ranges of x_{i1} and x_{i2} , where x_{i2} selected range is shifted to the right of $x_{i2} = 0$. Then, true β_0 is shifted to the right side of its AIEs. True β_1 location is always inside each AIE, where it is not strongly impacted by shifting values of x_{i2} . The situation is different for true β_2 which tends to move to the left side of its AIEs.

Table 4.2: AIEs with $x_1 = -0.5$ and $x_2 = 0.5$.

IDM						
$\nu = 1$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.015	0.016	0.031	0.352	0.433	0.081
$\nu = 2$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.015	0.017	0.032	0.347	0.447	0.1
IBLnM						
$\mu_j \in \{-0.35, -0.45\}, \mu_k \in \{0.35, 0.45\},$ $\sigma_j = \sigma_k \in \{0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.015	0.017	0.032	0.346	0.445	0.099
$\mu_j \in \{-0.5, -1\}, \mu_k \in \{0.5, 1\},$ $\sigma_j = \sigma_k \in \{0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.016	0.016	0.032	0.339	0.451	0.112
$\mu_j \in \{-0.35, -0.45\}, \mu_k \in \{0.35, 0.45\},$ $\sigma_j = \sigma_k \in \{1.0, 1.08, 1.16, 1.24, 1.32, 1.40, 1.48\}$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.015	0.015	0.030	0.354	0.434	0.080
$\mu_j \in \{-0.5, -1\}, \mu_k \in \{0.5, 1\},$ $\sigma_j = \sigma_k \in \{1.0, 1.08, 1.16, 1.24, 1.32, 1.40, 1.48\}$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.014	0.015	0.029	0.352	0.438	0.086

Table 4.3: AIEs with $x_1 = -1$ and $x_2 = 1$.

IDM						
$\nu = 1$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.014	0.014	0.028	0.361	0.43	0.069
$\nu = 2$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.015	0.015	0.030	0.354	0.445	0.091
IBLnM						
$\mu_j \in \{-0.35, -0.45\}, \mu_k \in \{0.35, 0.45\},$ $\sigma_j = \sigma_k \in \{0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.014	0.016	0.030	0.363	0.443	0.08
$\mu_j \in \{-0.5, -1\}, \mu_k \in \{0.5, 1\},$ $\sigma_j = \sigma_k \in \{0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.015	0.015	0.030	0.341	0.449	0.108
$\mu_j \in \{-0.35, -0.45\}, \mu_k \in \{0.35, 0.45\},$ $\sigma_j = \sigma_k \in \{1.0, 1.08, 1.16, 1.24, 1.32, 1.40, 1.48\}$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.014	0.015	0.029	0.363	0.427	0.064
$\mu_j \in \{-0.5, -1\}, \mu_k \in \{0.5, 1\},$ $\sigma_j = \sigma_k \in \{1.0, 1.08, 1.16, 1.24, 1.32, 1.40, 1.48\}$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.013	0.014	0.027	0.358	0.436	0.078

Table 4.4: AIEs with $x_1 = -2.5$ and $x_2 = 2.5$.

IDM						
$\nu = 1$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.013	0.014	0.027	0.389	0.422	0.033
$\nu = 2$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.015	0.015	0.030	0.382	0.442	0.060
IBLnM						
$\mu_j \in \{-0.35, -0.45\}, \mu_k \in \{0.35, 0.45\},$ $\sigma_j = \sigma_k \in \{0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.013	0.015	0.028	0.382	0.432	0.050
$\mu_j \in \{-0.5, -1\}, \mu_k \in \{0.5, 1\},$ $\sigma_j = \sigma_k \in \{0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.014	0.015	0.029	0.352	0.447	0.095
$\mu_j \in \{-0.35, -0.45\}, \mu_k \in \{0.35, 0.45\},$ $\sigma_j = \sigma_k \in \{1.0, 1.08, 1.16, 1.24, 1.32, 1.40, 1.48\}$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.014	0.014	0.028	0.378	0.411	0.033
$\mu_j \in \{-0.5, -1\}, \mu_k \in \{0.5, 1\},$ $\sigma_j = \sigma_k \in \{1.0, 1.08, 1.16, 1.24, 1.32, 1.40, 1.48\}$						
	$\hat{\beta}_0$	$\bar{\beta}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\beta}_1$	$\text{imp}(\hat{\beta}_1)$
$m = 2, n = 20$	-0.013	0.014	0.027	0.385	0.422	0.037

Table 4.5: Multiple covariate values under different ranges.

$x_i \in [-0.5, 0.5]$						
$m = 5$	$x_1 = -0.5$	$x_2 = -0.25$	$x_3 = 0$	$x_4 = 0.25$	$x_5 = 0.5$	
$m = 10$	$x_1 = -0.5$	$x_2 = -0.39$	$x_3 = -0.28$	$x_4 = -0.17$	$x_5 = -0.06$	$x_6 = 0.07$
	$x_7 = 0.18$	$x_8 = 0.29$	$x_9 = 0.4$	$x_{10} = 0.5$		
$m = 20$	$x_1 = -0.5$	$x_2 = -0.45$	$x_3 = -0.4$	$x_4 = -0.35$	$x_5 = -0.3$	$x_6 = -0.25$
	$x_7 = -0.2$	$x_8 = -0.15$	$x_9 = -0.1$	$x_{10} = -0.05$	$x_{11} = 0$	$x_{12} = 0.05$
	$x_{13} = 0.1$	$x_{14} = 0.15$	$x_{15} = 0.2$	$x_{16} = 0.25$	$x_{17} = 0.3$	$x_{18} = 0.35$
	$x_{19} = 0.42$	$x_{20} = 0.5$				
$x_i \in [-1, 1]$						
$m = 5$	$x_1 = -1$	$x_2 = -0.5$	$x_3 = 0$	$x_4 = 0.5$	$x_5 = 1$	
$m = 10$	$x_1 = -1$	$x_2 = -0.8$	$x_3 = -0.6$	$x_4 = -0.4$	$x_5 = -0.2$	$x_6 = 0.2$
	$x_7 = 0.4$	$x_8 = 0.6$	$x_9 = 0.8$	$x_{10} = 1$		
$m = 20$	$x_1 = -1$	$x_2 = -0.9$	$x_3 = -0.8$	$x_4 = -0.7$	$x_5 = -0.6$	$x_6 = -0.5$
	$x_7 = -0.4$	$x_8 = -0.3$	$x_9 = -0.2$	$x_{10} = -0.1$	$x_{11} = 0.1$	$x_{12} = 0.2$
	$x_{13} = 0.3$	$x_{14} = 0.4$	$x_{15} = 0.5$	$x_{16} = 0.6$	$x_{17} = 0.7$	$x_{18} = 0.8$
	$x_{19} = 0.9$	$x_{20} = 1$				
$x_i \in [-2.5, 2.5]$						
$m = 5$	$x_1 = -2.5$	$x_2 = -1$	$x_3 = 0$	$x_4 = 1$	$x_5 = 2.5$	
$m = 10$	$x_1 = -2.5$	$x_2 = -2$	$x_3 = -1.5$	$x_4 = -1$	$x_5 = -0.5$	$x_6 = 0.5$
	$x_7 = 1$	$x_8 = 1.5$	$x_9 = 2$	$x_{10} = 2.5$		
$m = 20$	$x_1 = -2.5$	$x_2 = -2.25$	$x_3 = -2$	$x_4 = -1.75$	$x_5 = -1.5$	$x_6 = -1.25$
	$x_7 = -1$	$x_8 = -0.75$	$x_9 = -0.5$	$x_{10} = -0.25$	$x_{11} = 0.25$	$x_{12} = 0.5$
	$x_{13} = 0.75$	$x_{14} = 0.1$	$x_{15} = 1.25$	$x_{16} = 1.5$	$x_{17} = 1.75$	$x_{18} = 2$
	$x_{19} = 2.25$	$x_{20} = 2.5$				

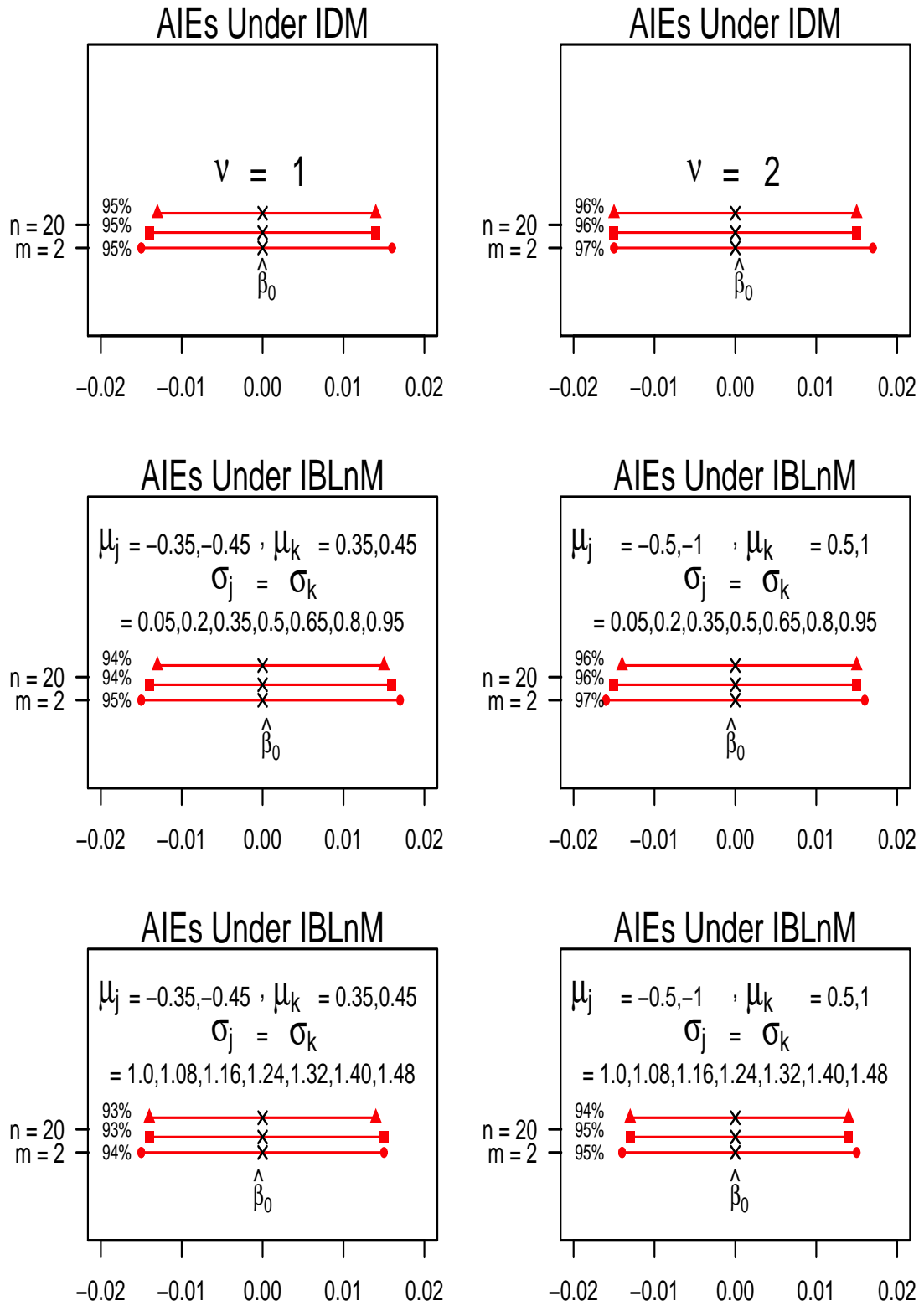


Figure 4.1: Plots of AIEs of β_0 ended by \bullet when $x_1 = -0.5$ and $x_2 = 0.5$, \blacksquare when $x_1 = -1$ and $x_2 = 1$ and \blacktriangle when $x_1 = -2.5$ and $x_2 = 2.5$. Percentage values are for single interval estimates that include true β_0 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

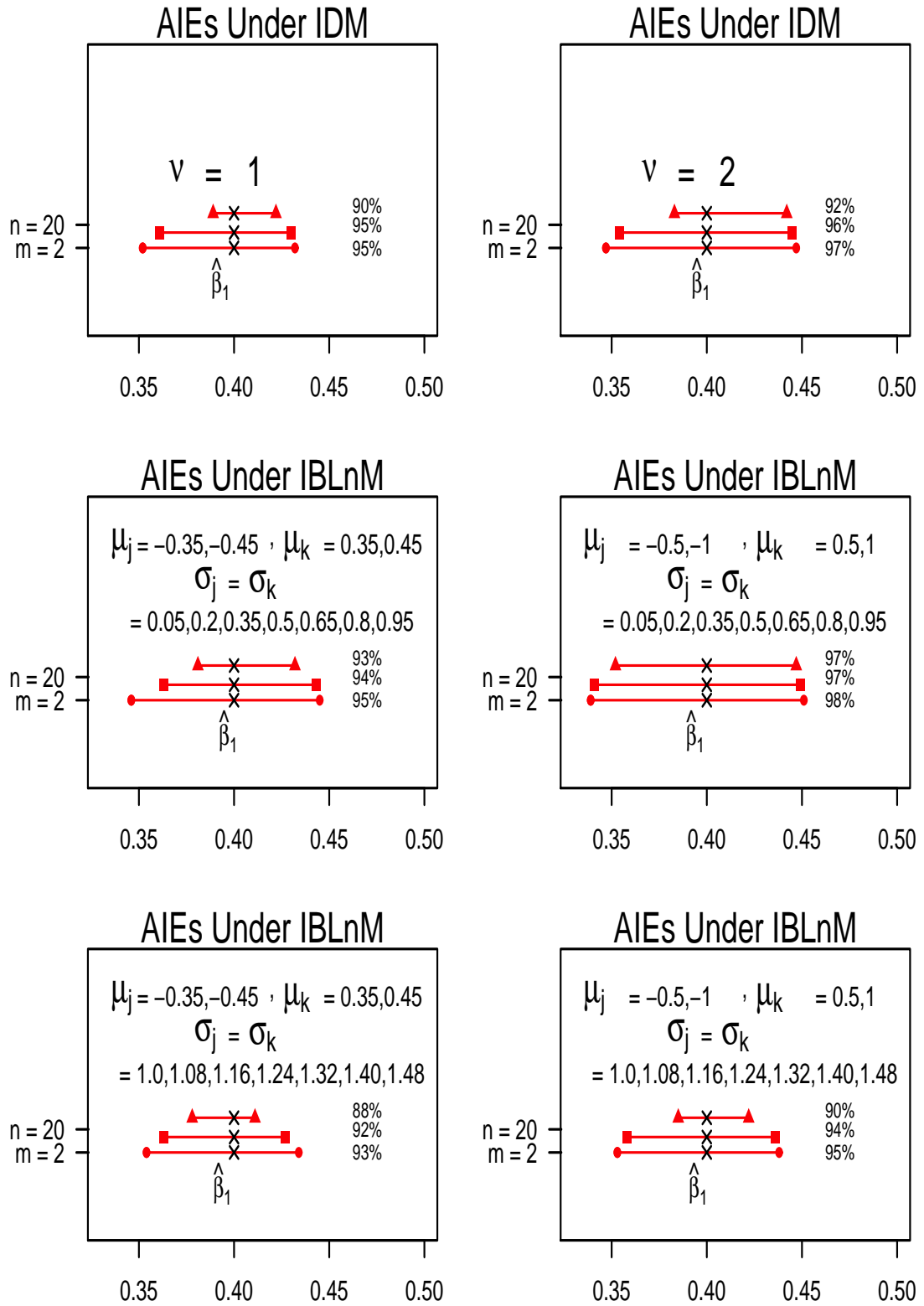


Figure 4.2: Plots of AIEs of β_1 ended by \bullet when $x_1 = -0.5$ and $x_2 = 0.5$, \blacksquare when $x_1 = -1$ and $x_2 = 1$ and \blacktriangle when $x_1 = -2.5$ and $x_2 = 2.5$. Percentage values are for single interval estimates that include true β_1 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

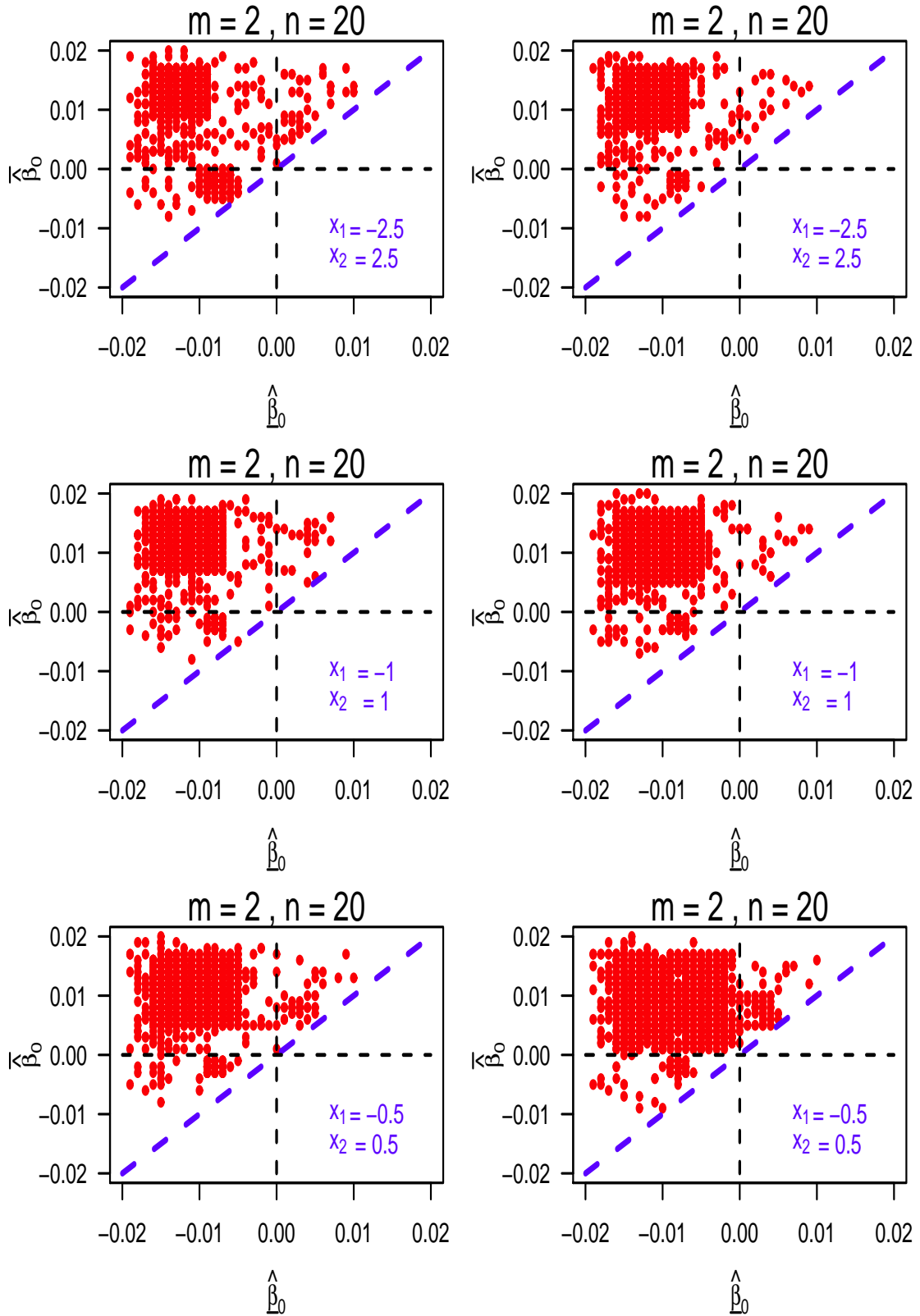


Figure 4.3: Left-column: Plots of lower vs upper estimates of β_0 under symmetric two x_i 's design with $\nu = 1$, $m = 2$ and $n_1 = n_2 = 20$, that correspond to the AIEs in the top-left plot of Figure 4.1. Right-column: For the top-right plot of Figure 4.1 with $\nu = 2$, $m = 2$ and $n_1 = n_2 = 20$.

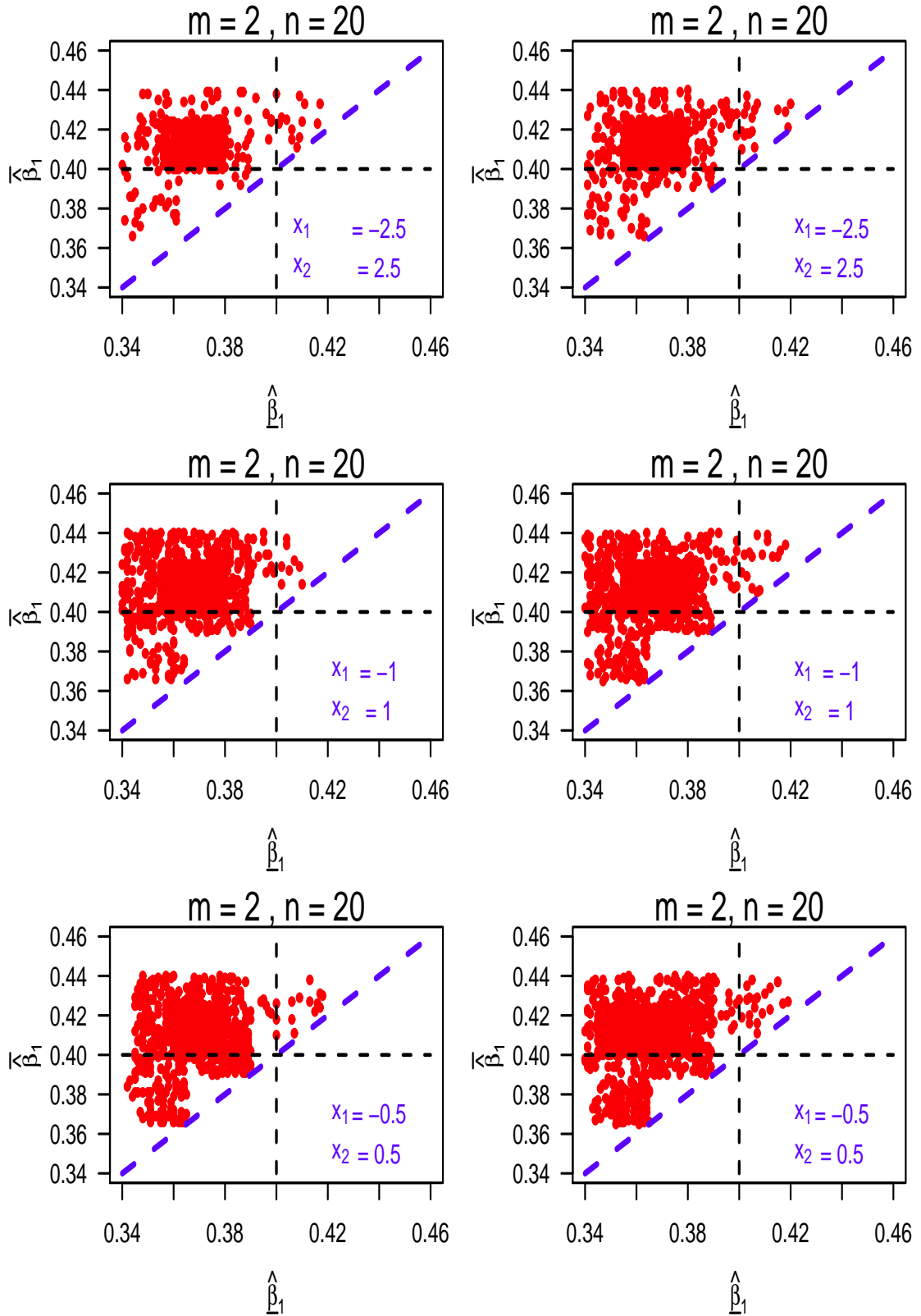


Figure 4.4: Left-column: Plots of lower vs upper estimates of β_1 under symmetric two x_i 's design with $\nu = 1$, $m = 2$ and $n_1 = n_2 = 20$, that correspond to AIEs in the top-left plot of Figure 4.2. Right-column: For the top-right plot of Figure 4.2 with $\nu = 2$, $m = 2$ and $n_1 = n_2 = 20$.

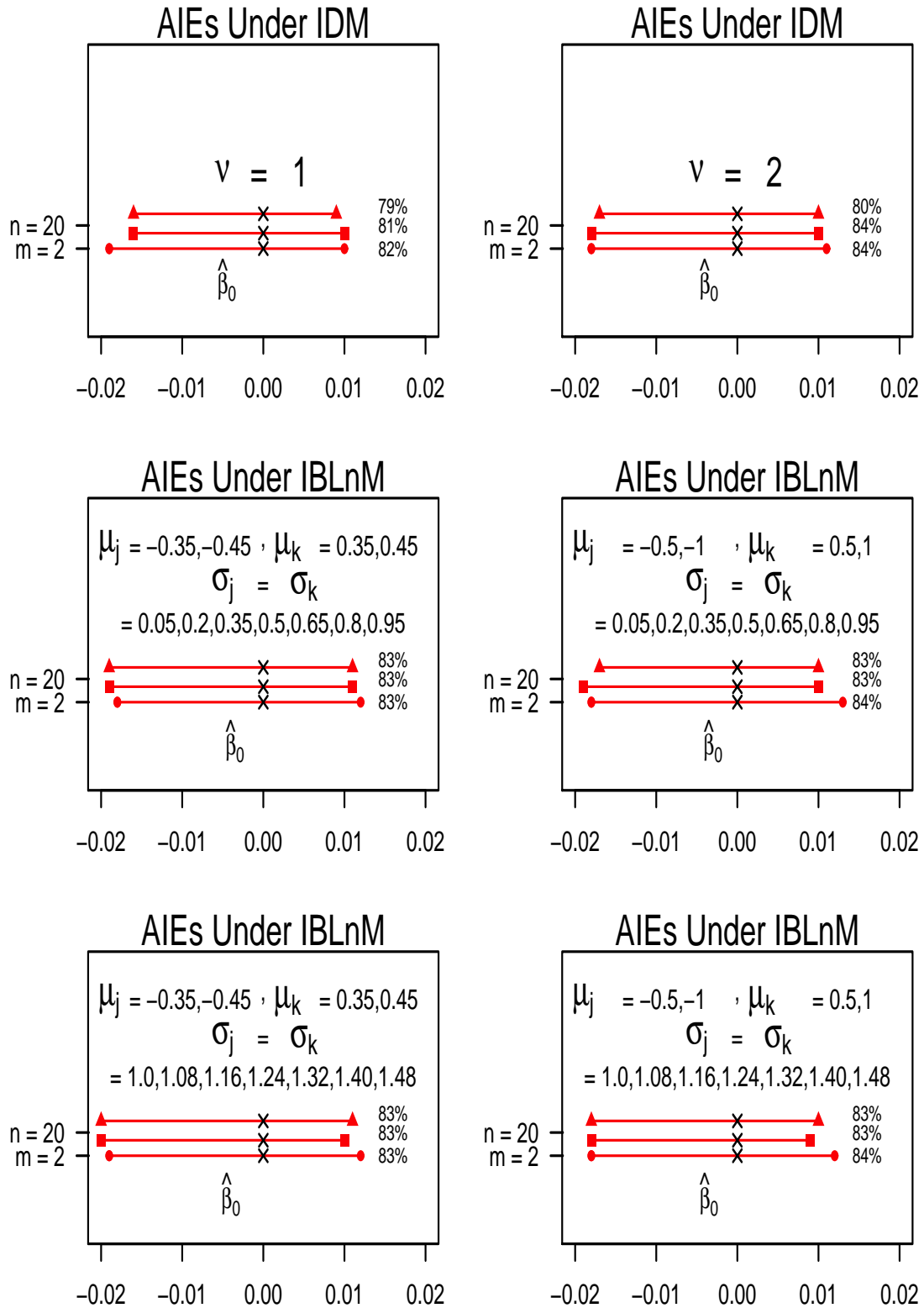


Figure 4.5: Plots of AIEs of β_0 ended by \bullet when $x_1 = 1$ and $x_2 = 2$, \blacksquare when $x_1 = 1$ and $x_2 = 3$ and \blacktriangle when $x_1 = 1$ and $x_2 = 4$. Percentage values are for single interval estimates that include true β_0 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

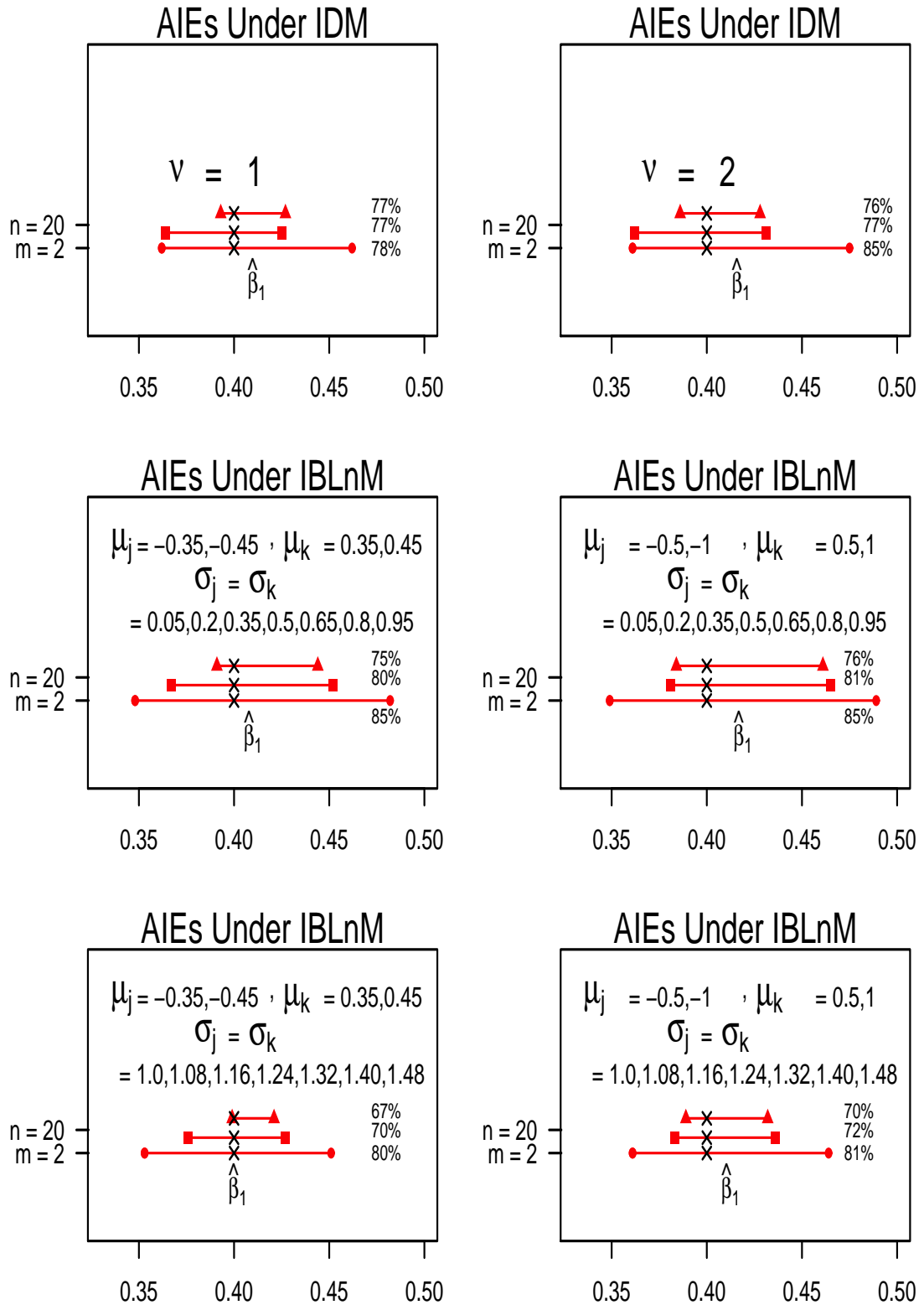


Figure 4.6: Plots of AIEs of β_1 ended by \bullet when $x_1 = 1$ and $x_2 = 2$, \blacksquare when $x_1 = 1$ and $x_2 = 3$ and \blacktriangle when $x_1 = 1$ and $x_2 = 4$. Percentage values are for single interval estimates that include true β_1 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

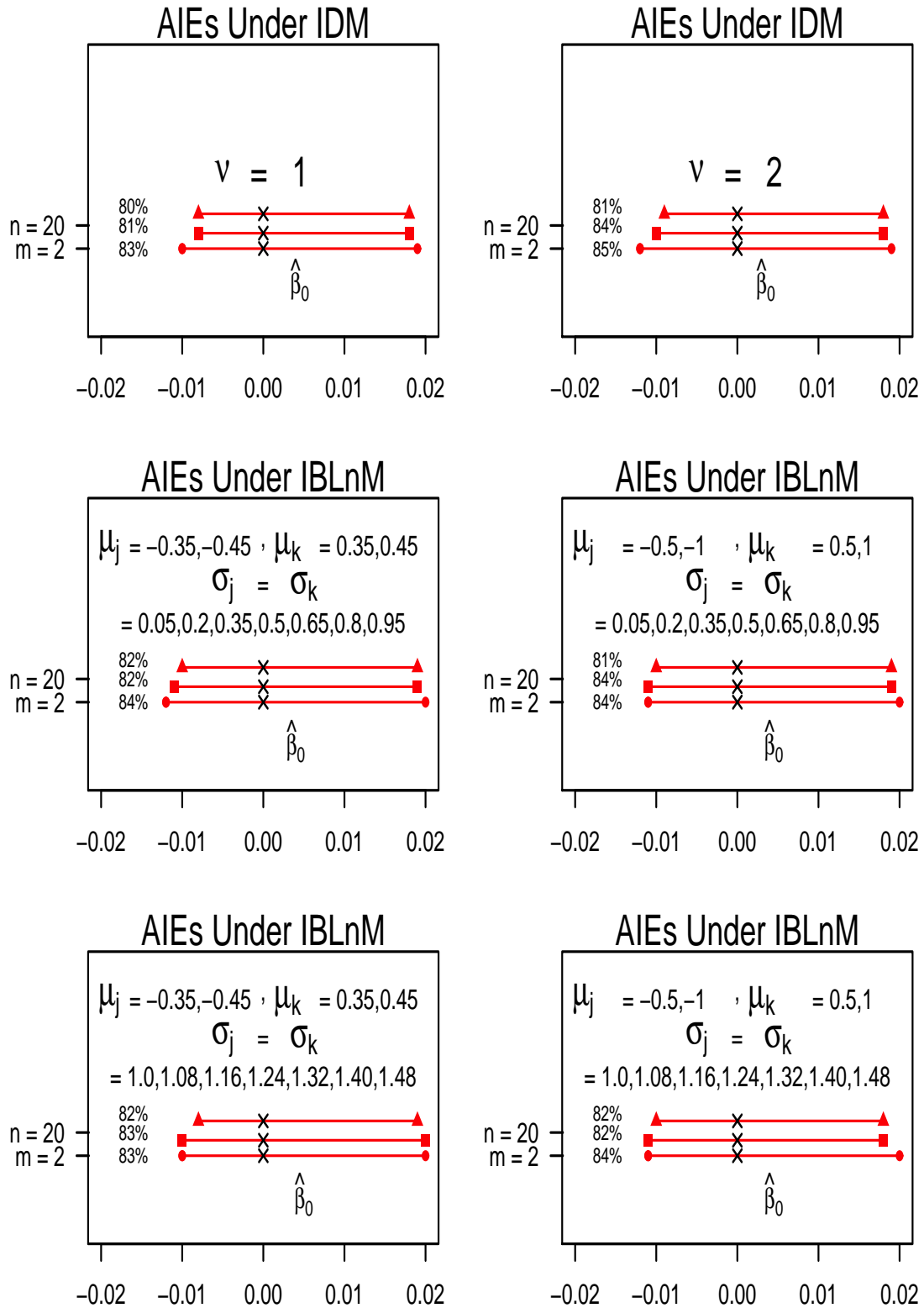


Figure 4.7: Plots of AIEs of β_0 ended by \bullet when $x_1 = -2$ and $x_2 = -1$, \blacksquare when $x_1 = -3$ and $x_2 = -1$ and \blacktriangle when $x_1 = -4$ and $x_2 = -1$. Percentage values are for single interval estimates that include true β_0 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

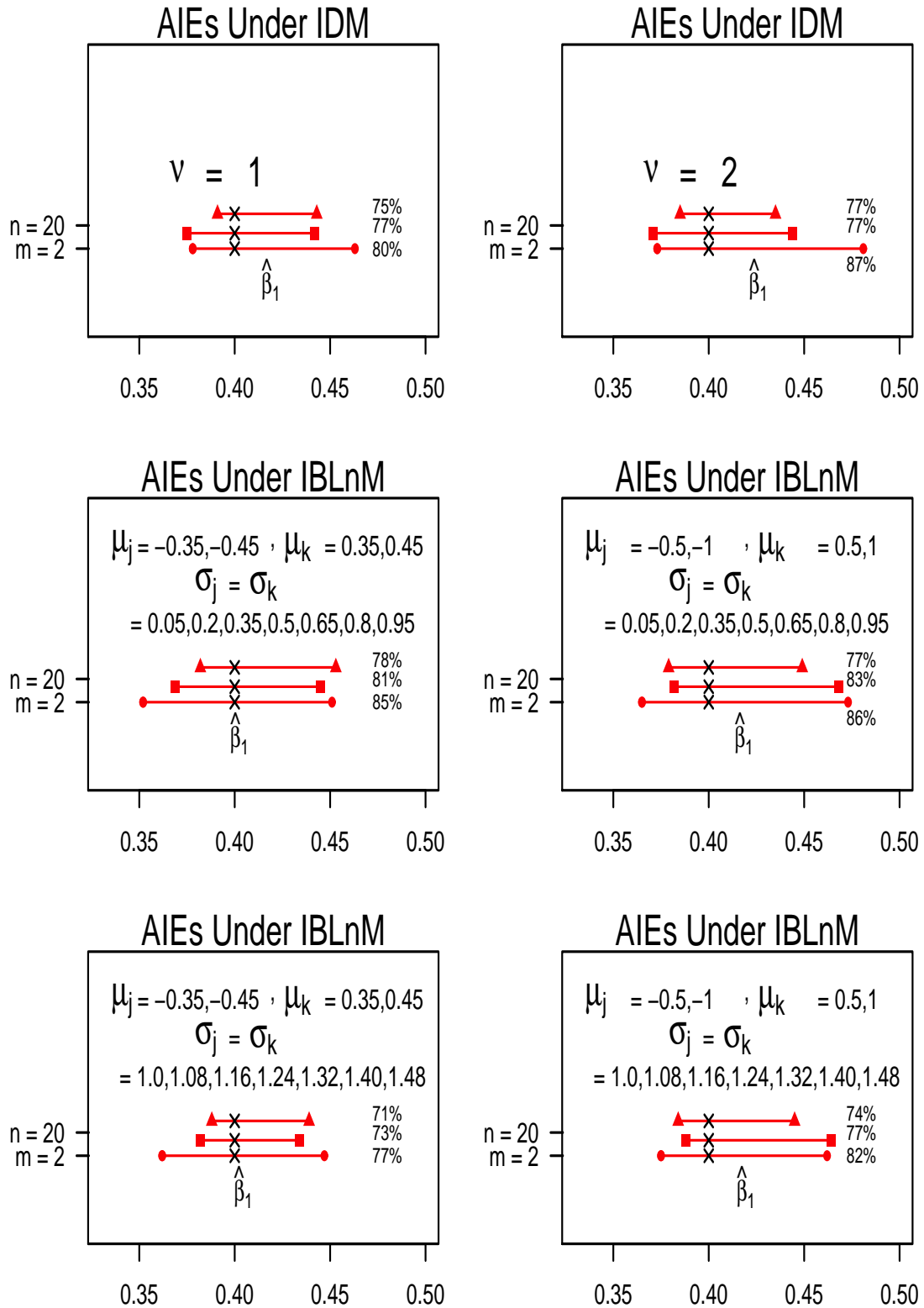


Figure 4.8: Plots of AIEs of β_1 ended by \bullet when $x_1 = -2$ and $x_2 = -1$, \blacksquare when $x_1 = -3$ and $x_2 = -1$ and \blacktriangle when $x_1 = -4$ and $x_2 = -1$. Percentage values are for single interval estimates that include true β_1 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

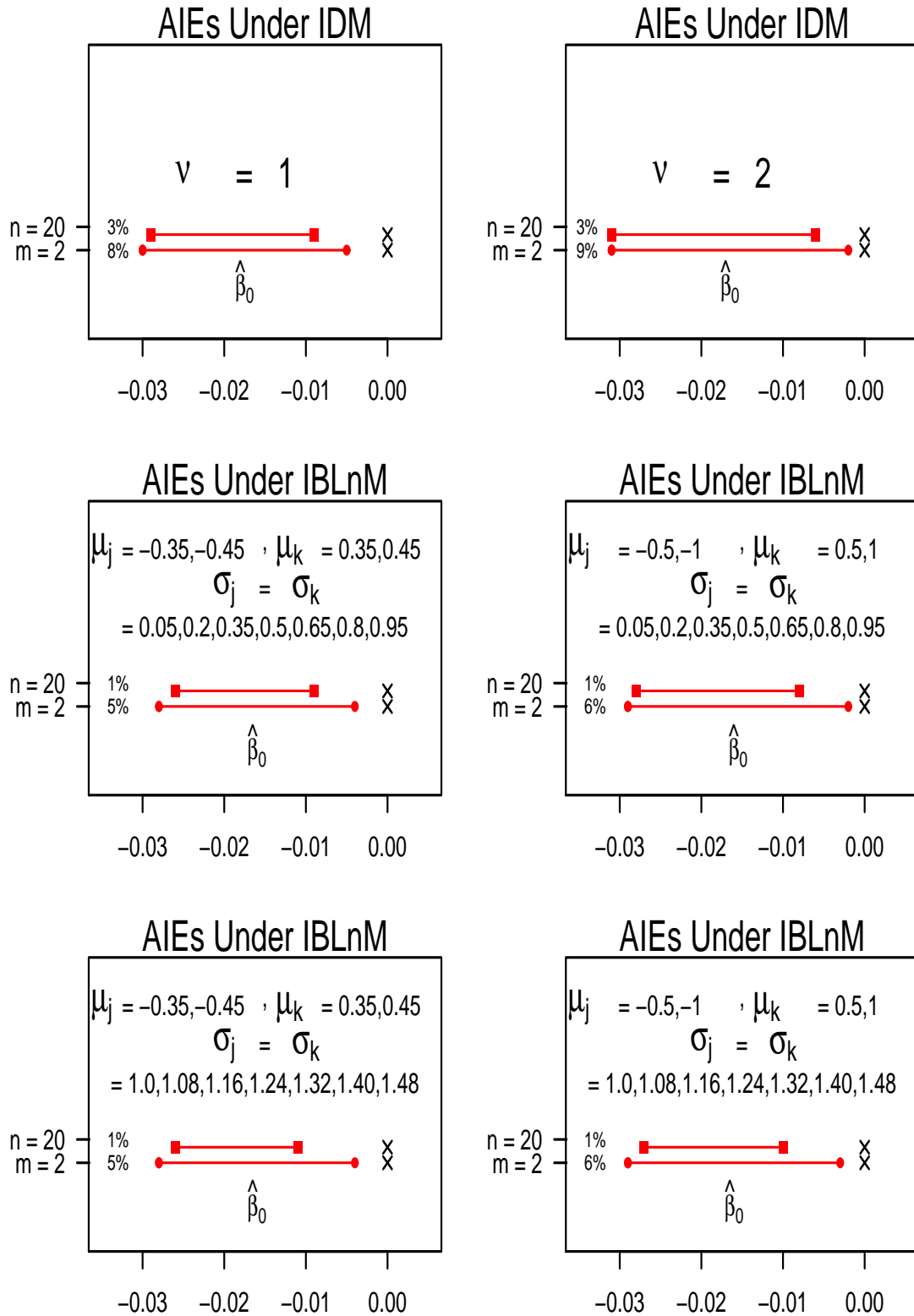


Figure 4.9: Plots of AIEs of β_0 ended by \bullet when $x_1 = 3.5$ and $x_2 = 4.5$, \blacksquare when $x_1 = 3$ and $x_2 = 5$. Percentage values are for single interval estimates that include true β_0 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

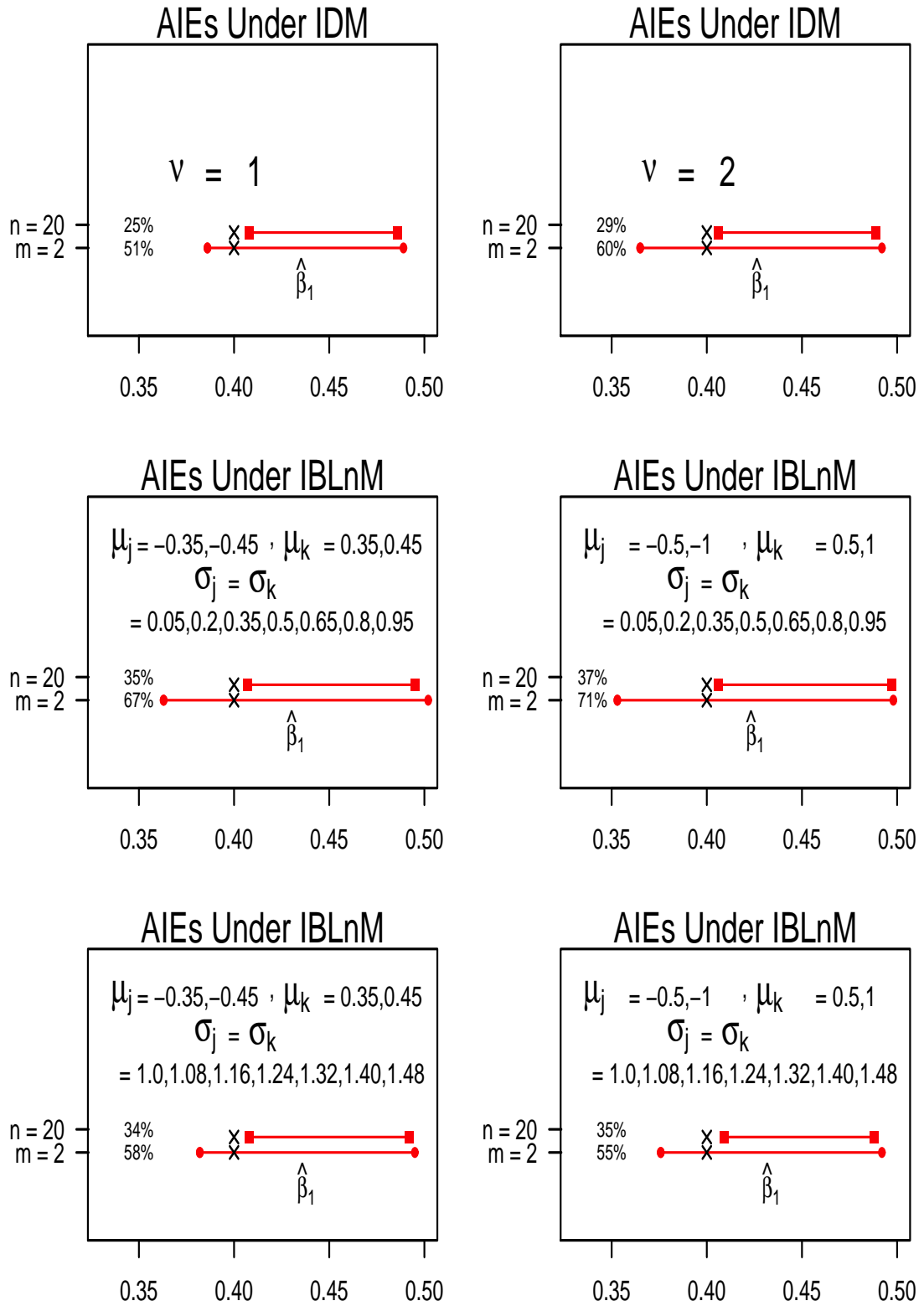


Figure 4.10: Plots of AIEs of β_1 ended by \bullet when $x_1 = 3.5$ and $x_2 = 4.5$, \blacksquare when $x_1 = 3$ and $x_2 = 5$. Percentage values are for single interval estimates that include true β_1 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

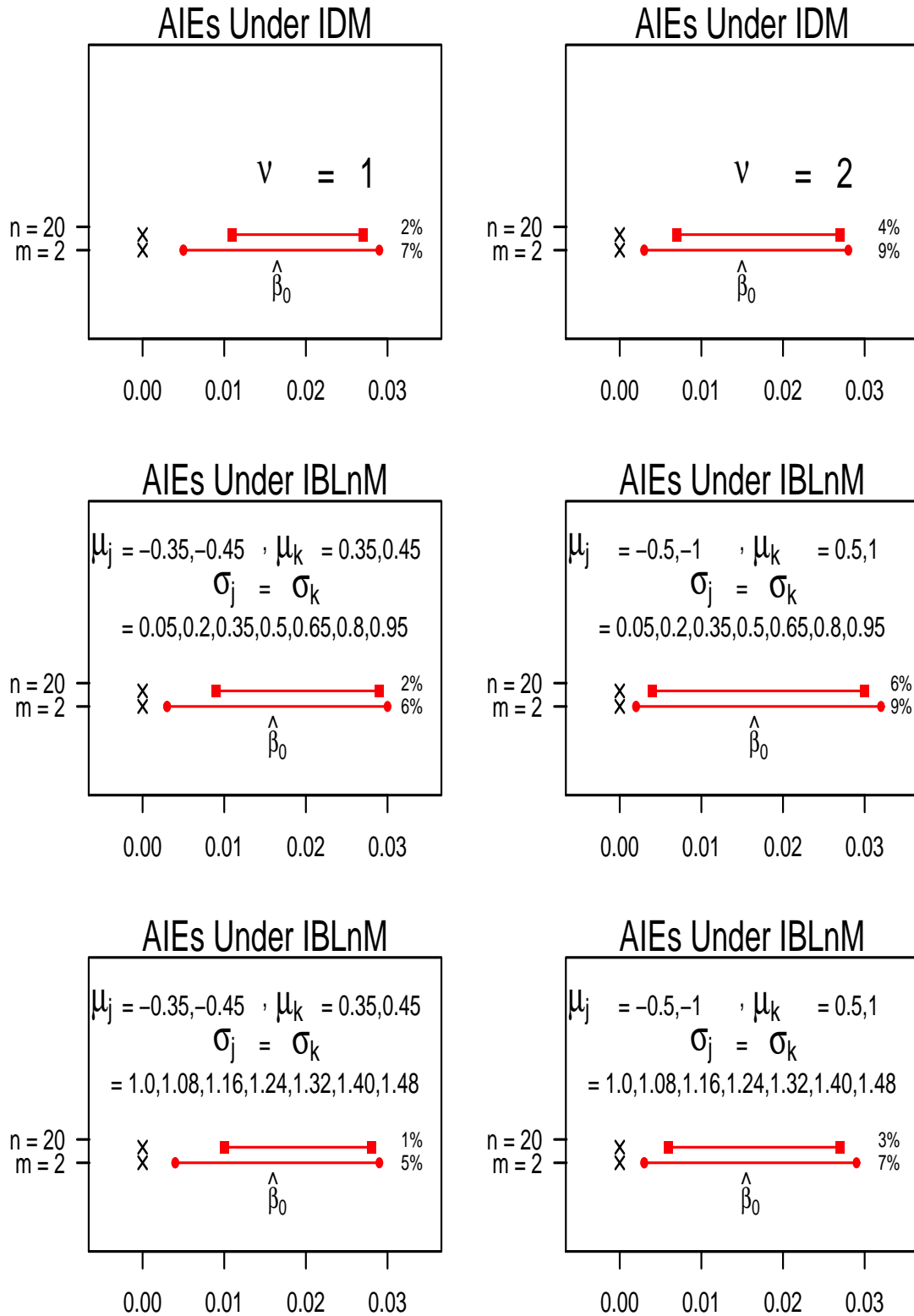


Figure 4.11: Plots of AIEs of β_0 ended by \bullet when $x_1 = -4.5$ and $x_2 = -3.5$, \blacksquare when $x_1 = -5$ and $x_2 = -3$. Percentage values are for single interval estimates that include true β_0 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

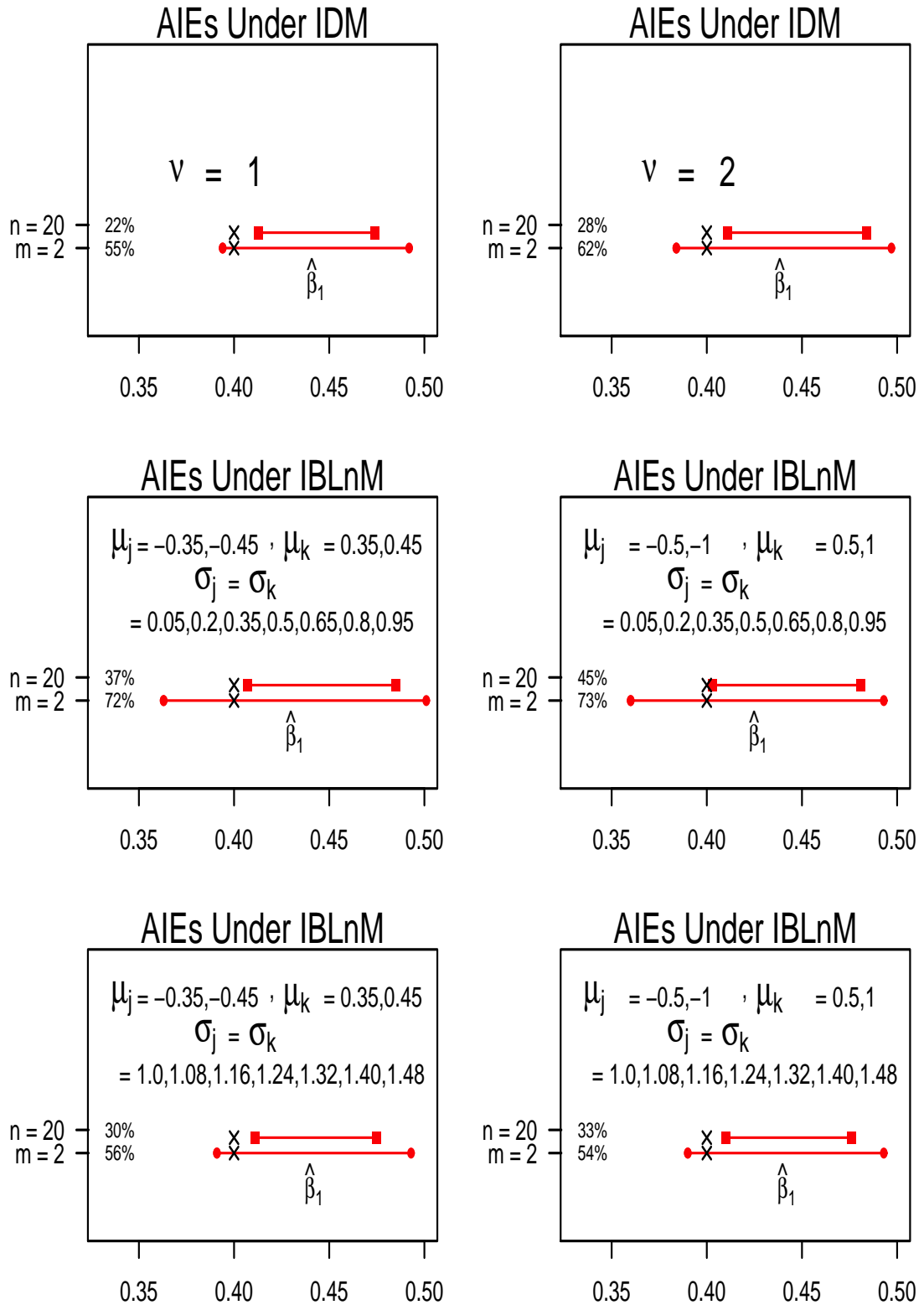


Figure 4.12: Plots of AIEs of β_1 ended by \bullet when $x_1 = -4.5$ and $x_2 = -3.5$, \blacksquare when $x_1 = -5$ and $x_2 = -3$. Percentage values are for single interval estimates that include true β_1 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

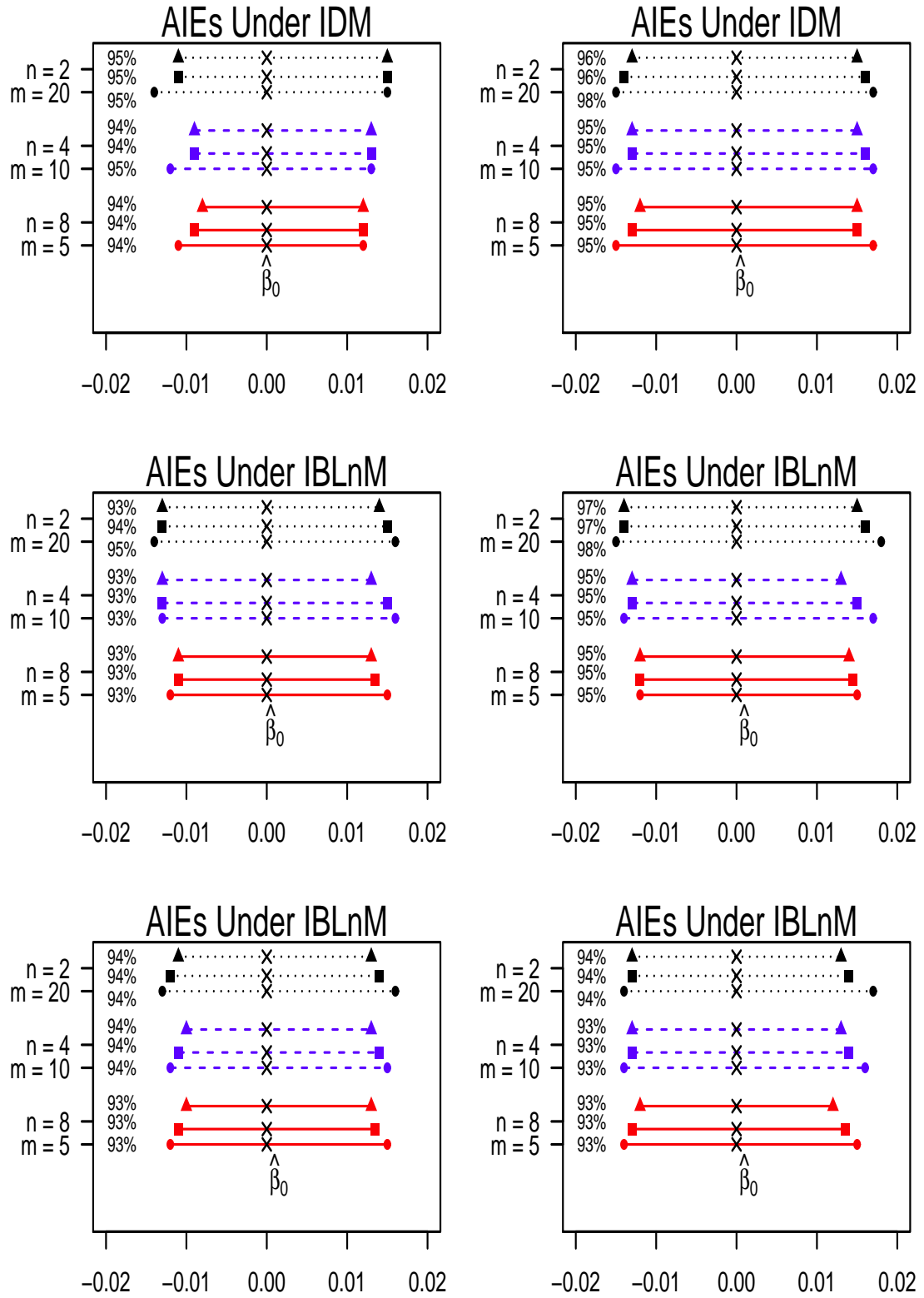


Figure 4.13: Plots of AIEs of β_0 ended by \bullet when $x_i \in [-0.5, 0.5]$, $i = 1, \dots, m$, \blacksquare when $x_i \in [-1, 1]$ and \blacktriangle when $x_i \in [-2.5, 2.5]$. Percentage values are for single interval estimates that include true β_0 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

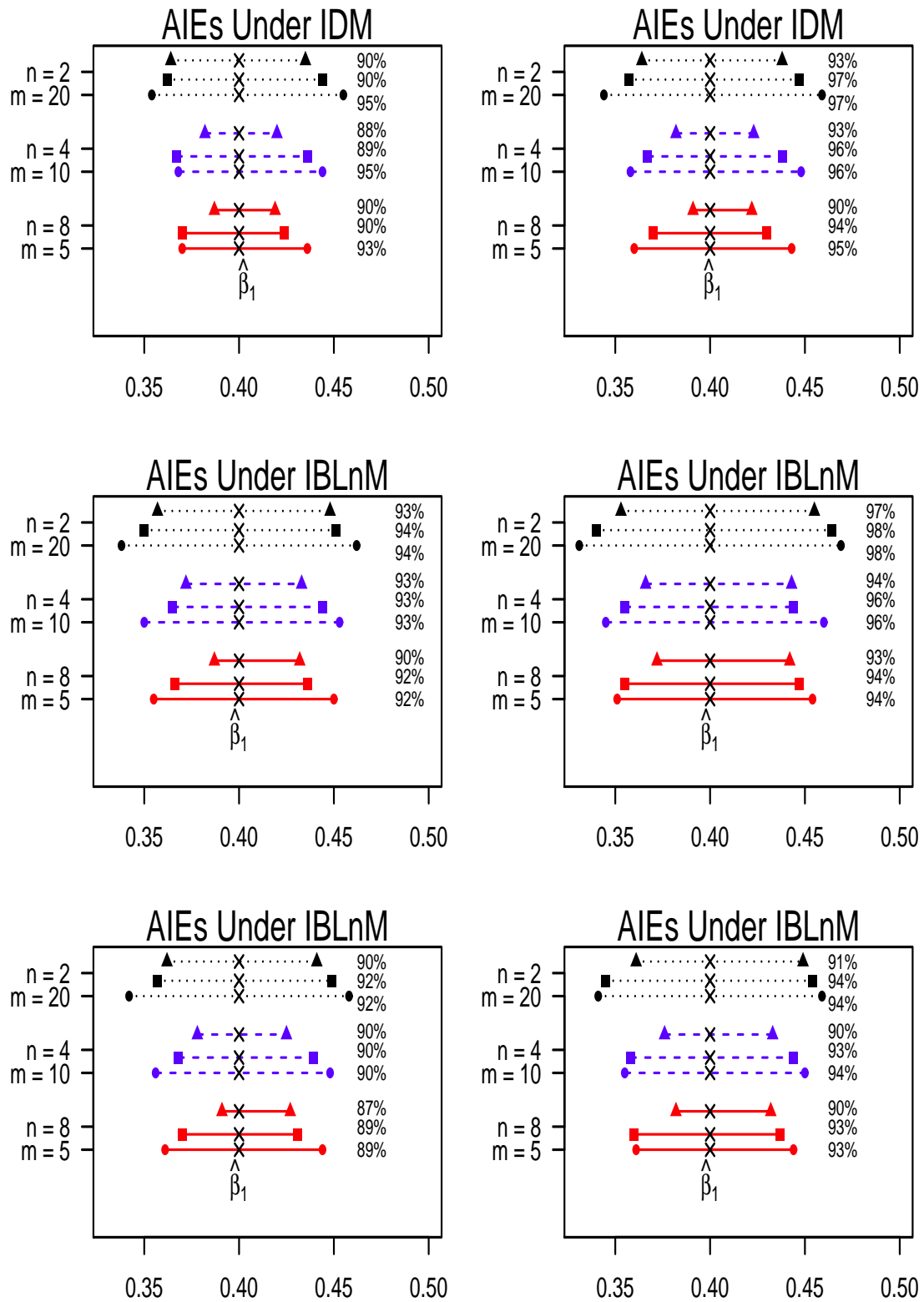


Figure 4.14: Plots of AIEs of β_1 ended by \bullet when $x_i \in [-0.5, 0.5]$, $i = 1, \dots, m$, \blacksquare when $x_i \in [-1, 1]$ and \blacktriangle when $x_i \in [-2.5, 2.5]$. Percentage values are for single interval estimates that include true β_1 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

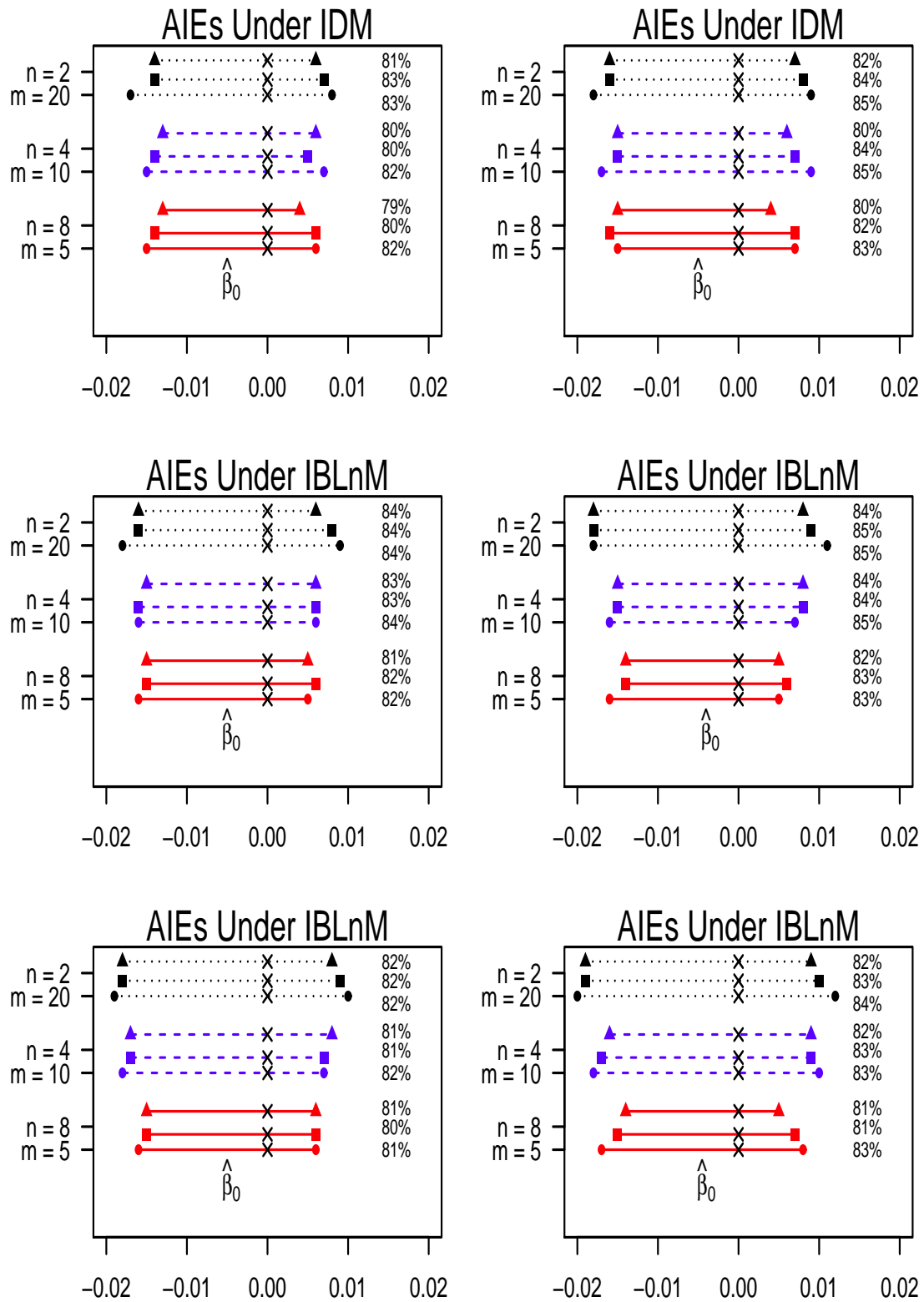


Figure 4.15: Plots of AIEs of β_0 ended by \bullet when $x_i \in [1, 2]$, $i = 1, \dots, m$, \blacksquare when $x_i \in [1, 3]$ and \blacktriangle when $x_i \in [1, 4]$. Percentage values are for single interval estimates that include true β_0 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

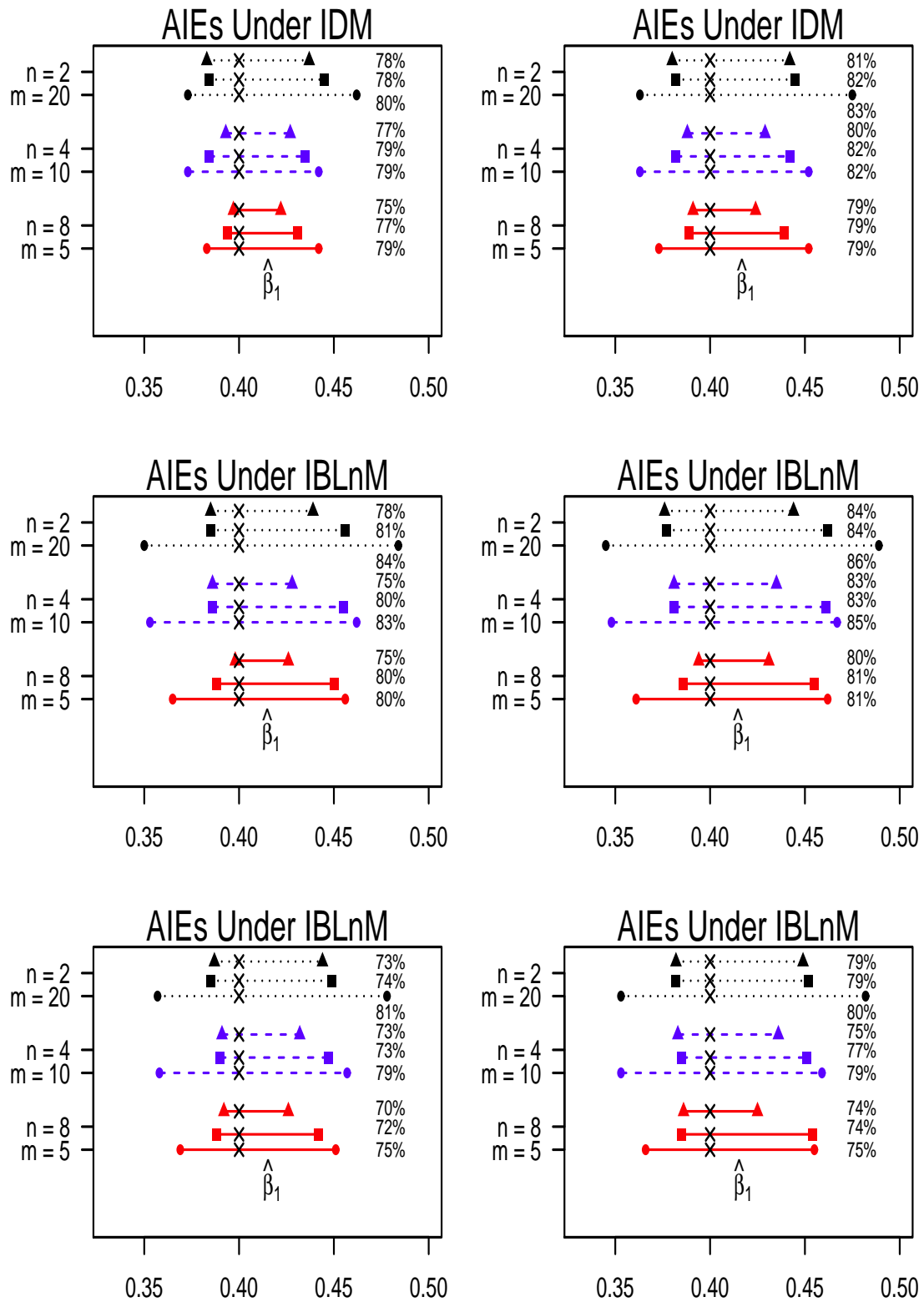


Figure 4.16: Plots of AIEs of β_1 ended by \bullet when $x_i \in [1, 2]$, $i = 1, \dots, m$, \blacksquare when $x_i \in [1, 3]$ and \blacktriangle when $x_i \in [1, 4]$. Percentage values are for single interval estimates that include true β_1 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

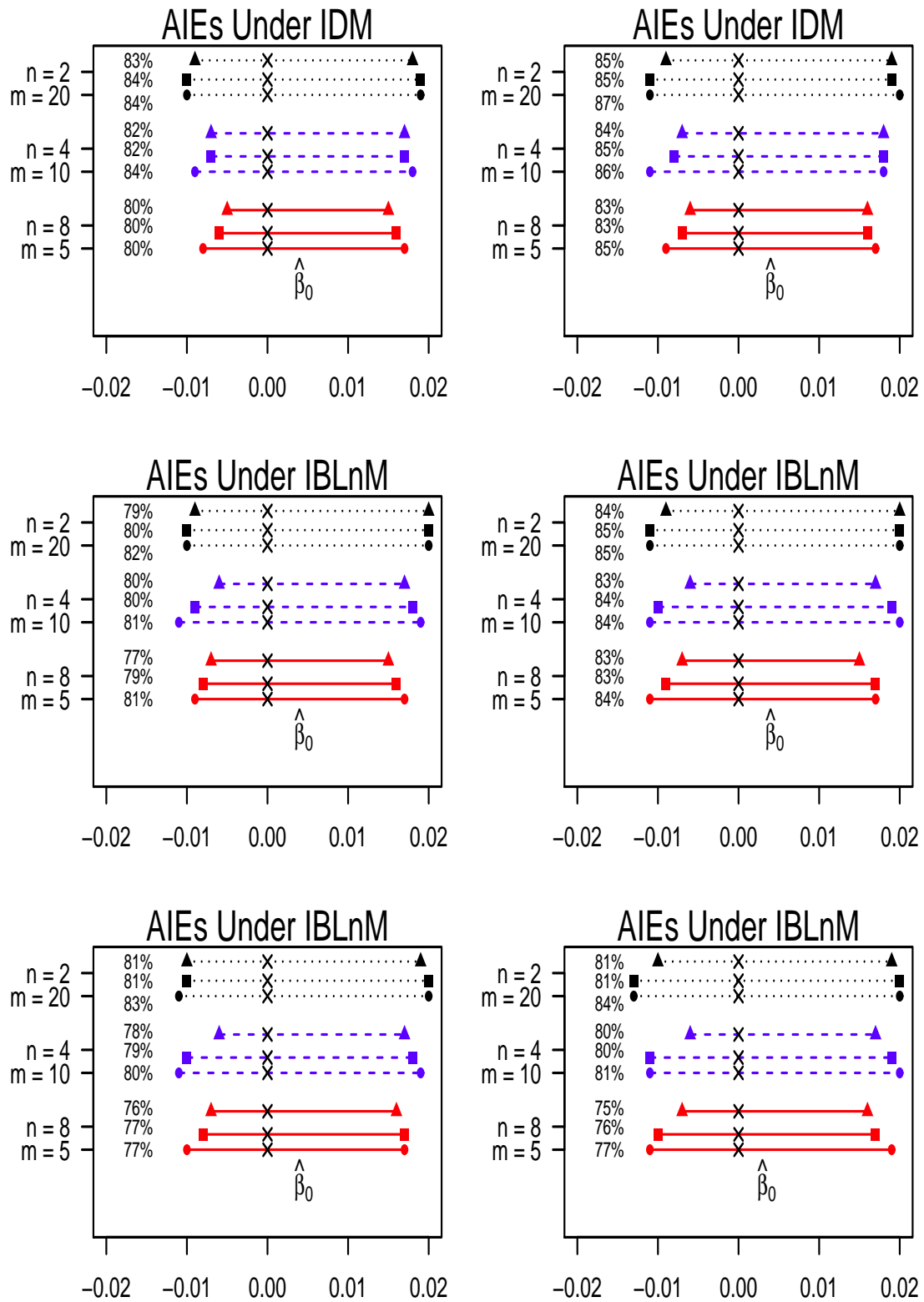


Figure 4.17: Plots of AIEs of β_0 ended by \bullet when $x_i \in [-2, -1]$, $i = 1, \dots, m$, \blacksquare when $x_i \in [-3, -1]$ and \blacktriangle when $x_i \in [-4, -1]$. Percentage values are for single interval estimates that include true β_0 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

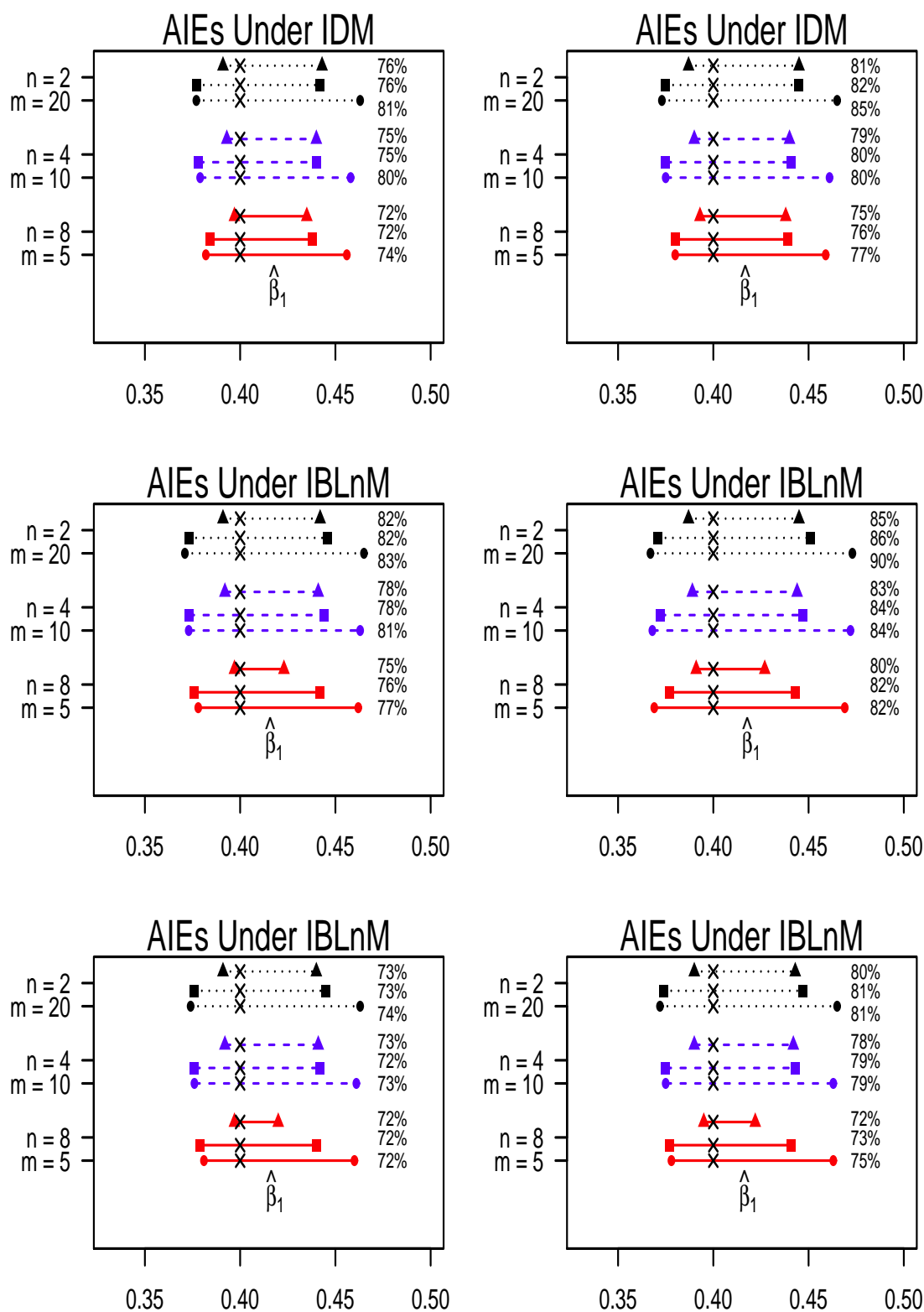


Figure 4.18: Plots of AIEs of β_1 ended by \bullet when $x_i \in [-2, -1]$, $i = 1, \dots, m$, \blacksquare when $x_i \in [-3, -1]$ and \blacktriangle when $x_i \in [-4, -1]$. Percentage values are for single interval estimates that include true β_1 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

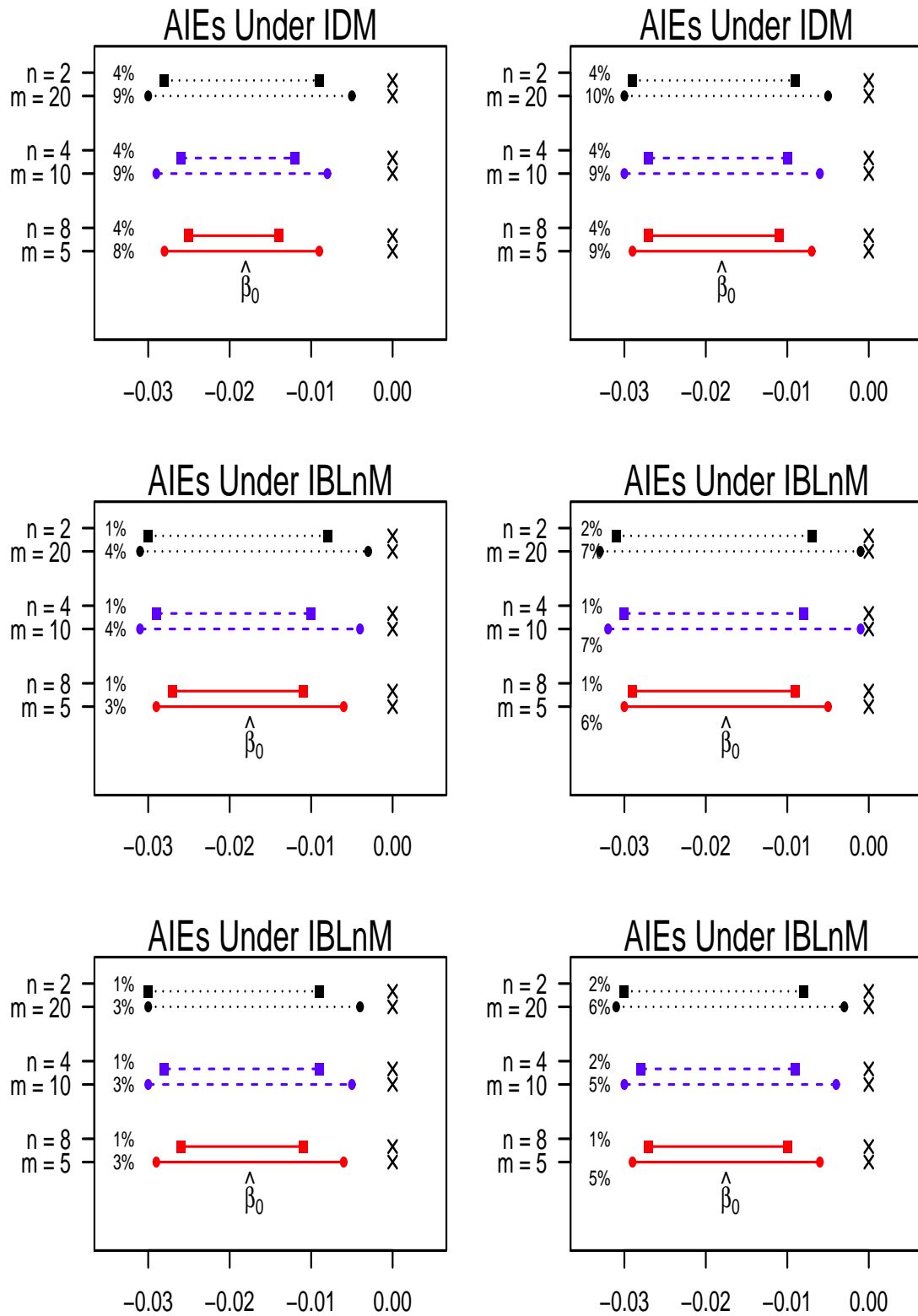


Figure 4.19: Plots of AIEs of β_0 ended by \bullet when $x_i \in [3.5, 4.5]$, $i = 1, \dots, m$, \blacksquare when $x_i \in [3, 5]$. Percentage values are for single interval estimates that include true β_0 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

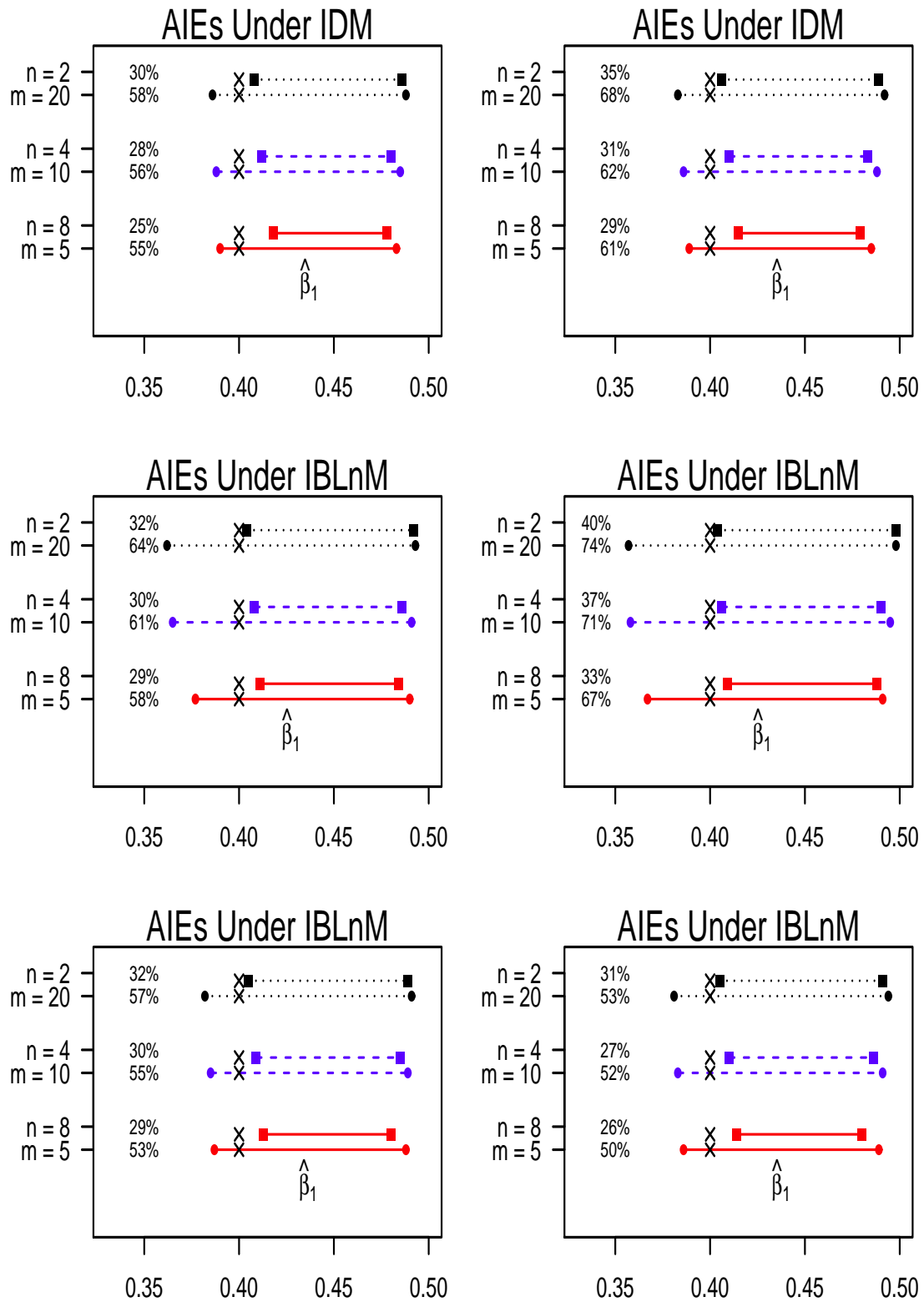


Figure 4.20: Plots of AIEs of β_1 ended by \bullet when $x_i \in [3.5, 4.5]$, $i = 1, \dots, m$, \blacksquare when $x_i \in [3, 5]$. Percentage values are for single interval estimates that include true β_1 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

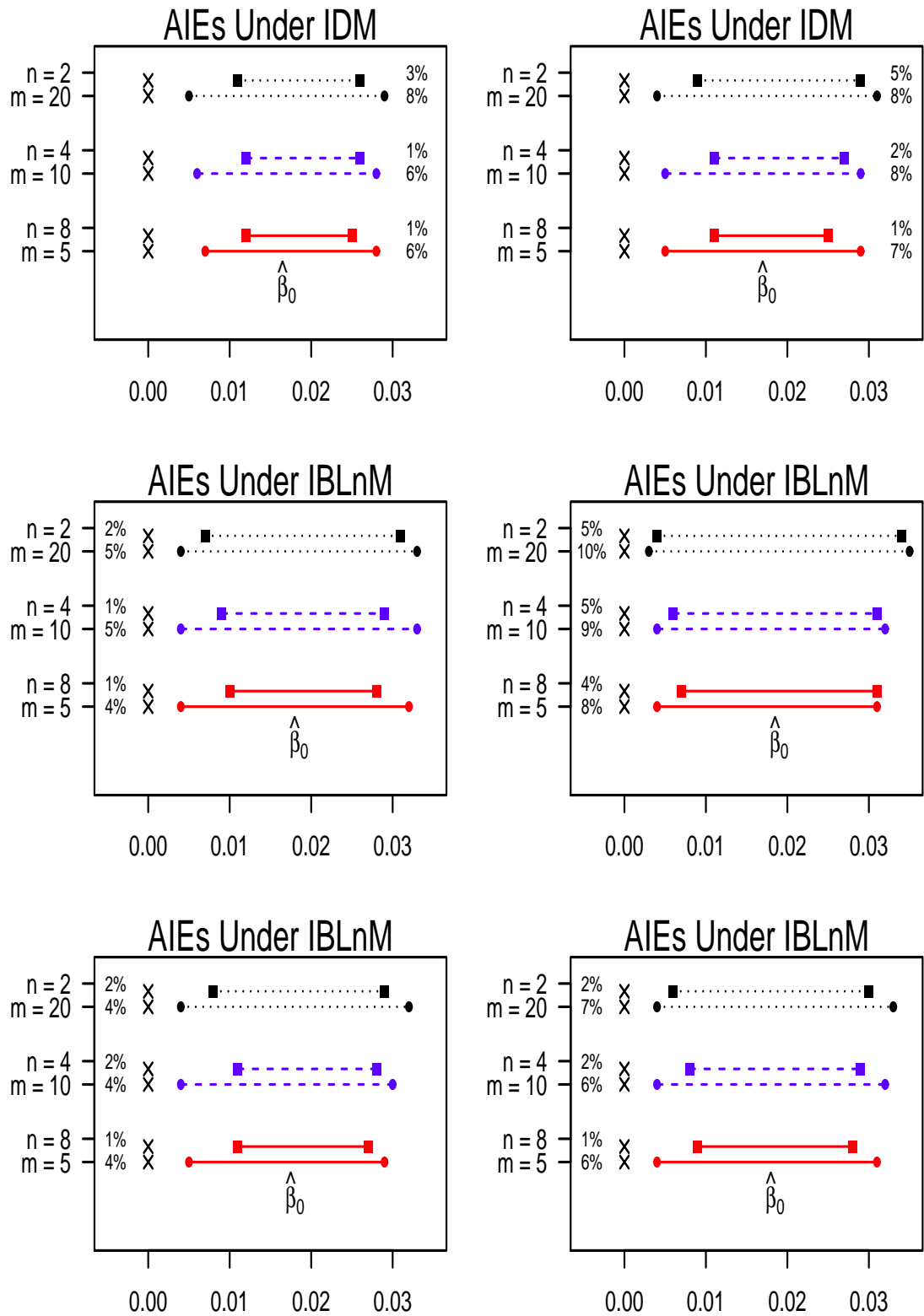


Figure 4.21: Plots of AIEs of β_0 ended by \bullet when $x_i \in [-4.5, -3.5]$, $i = 1, \dots, m$, \blacksquare when $x_i \in [-5, -3]$. Percentage values are for single interval estimates that include true β_0 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

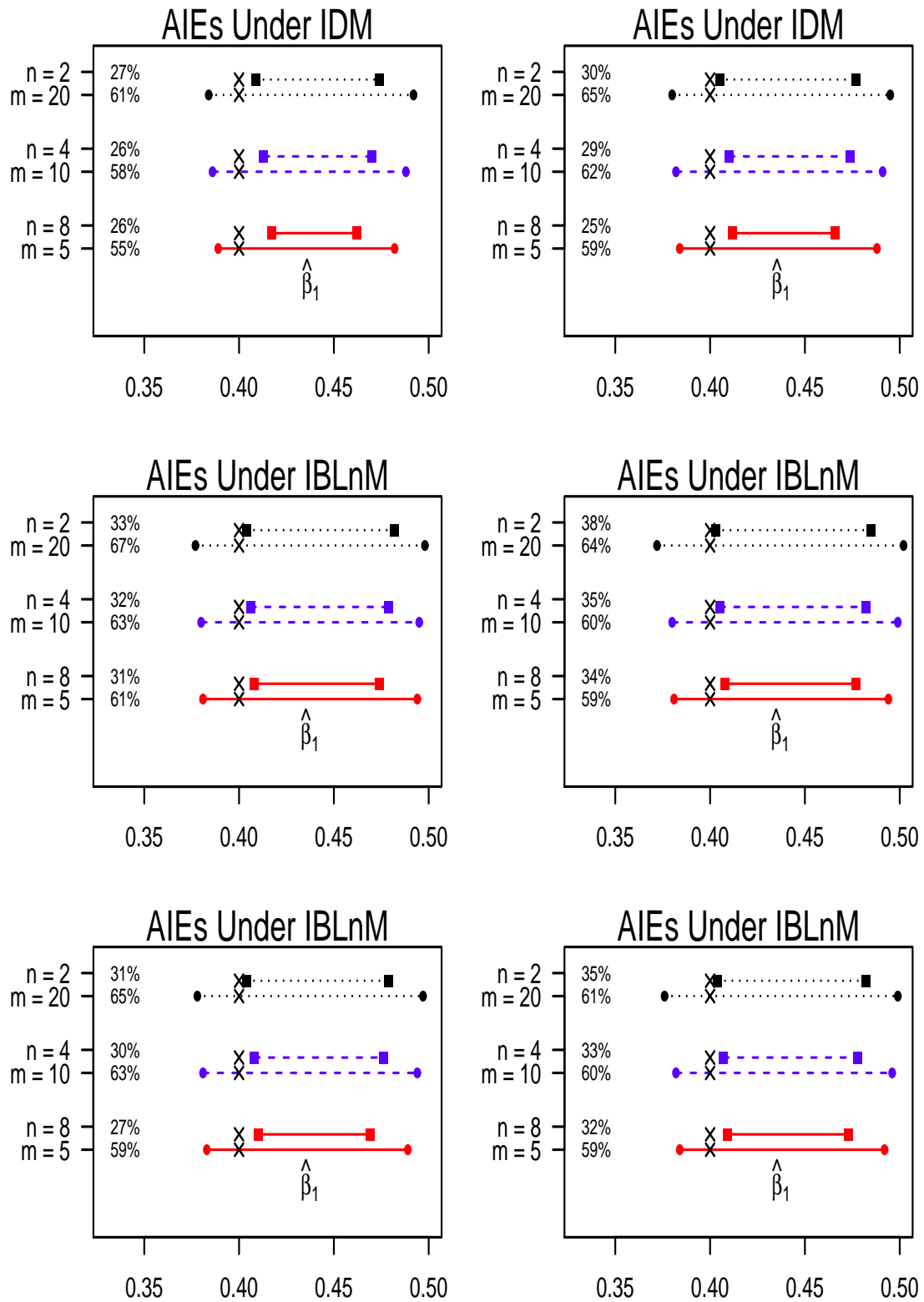


Figure 4.22: Plots of AIEs of β_1 ended by \bullet when $x_i \in [-4.5, -3.5]$, $i = 1, \dots, m$, \blacksquare when $x_i \in [-5, -3]$. Percentage values are for single interval estimates that include true β_1 . Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

CHAPTER 5

CONCLUSION AND FUTURE PLAN

This chapter presents a conclusion about 2-parameter ImpLogit model. The conclusion, presented in Section 5.1, will give recommendations for reducing imprecision of regression parameters estimates by describing a recommended design for 2-parameter ImpLogit model. Future research plan is given in Section 5.2.

5.1 Conclusion

The first step in building the imprecise logistic regression model is made. ImpLogit model applies probabilistic imprecision to regression parameters. ImpLogit model is a new imprecise logistic regression model that produces interval estimates for regression parameters.

In Chapter 1, the concept of probabilistic imprecision was presented. This required to have a quick look at the robust and hierarchical Bayesian methods. A general look at the historical development of Bayesian methods was given to understand circumstances that led statisticians to adopt the imprecise probabilistic approach. Chapter 1 introduced the imprecise Dirichlet model (IDM). Examples of applications of imprecise probability models were provided.

In Chapter 2, the successive differences of probabilities of occurrence in logistic regression model were assumed to follow the Dirichlet distribution. A new method for inducing a prior distribution to regression parameters in 2-parameter logistic regression model was given. This required to use the aggregation property in Dirichlet distribution to find the joint distribution for any selected θ_j and θ_k . The new method was applied to both cases of increasing and decreasing logistic regression models. The logit-normal distribution and its multivariate form were introduced. The bivariate logit-normal distribution was fitted

by inducing a prior distribution for regression parameters, subject to the new method. Prior and posterior densities in univariate and multivariate logit-normal distributions were shown, in appendices B and C, to belong to an exponential family of distributions.

In Chapter 3, the shape of 2-parameter ImpLogit model was studied in the first section. It was found that logistic functions under lower and upper values of β_0 and β_1 intersect at a negative covariate value. Also, lower and upper values of β_0 and β_1 do not always correspond to lower and upper logistic functions. The correspondence occurs over specific parts of the covariate values as being shown in Section 3.1. Chapter 3 presented the relation between $\boldsymbol{\beta} = [\beta_0, \beta_1]'$ and $\boldsymbol{\theta} = [\theta_j, \theta_k]'$, for any j, k . This meant to show that lower and upper $\boldsymbol{\beta} = [\beta_0, \beta_1]'$ do not necessarily correspond to lower and upper $\boldsymbol{\theta} = [\theta_j, \theta_k]'$.

In Chapter 3, the 2-parameter ImpLogit model was established, developed and fitted. A new imprecise probability model, called the imprecise bivariate logit-normal model (IBLnM), was built. The 2-parameter ImpLogit model was fitted under IDM and IBLnM. Having a smaller sample size was found to produce higher imprecision in regression parameters estimates.

An advanced simulation study for 2-parameter ImpLogit model was given in Chapter 4. The simulation study was made under several designs of 2-parameter ImpLogit model. The simulation study showed that different factors can associate to increasing (or decreasing) interval estimates lengths in regression parameters. The factors include changes in prior parameters values, locations and spread of fixed covariate values and the number of covariate values (m).

In the 2-parameter ImpLogit model, changes in ranges of ν in IDM, μ_i and σ_i in IBLnM and fixed covariate x_i values, $\forall i = 1, \dots, m$, affect the lengths of the produced interval estimates of regression parameters. This motivates to think in establishing a better and recommended design. The simulation results in Chapter 4 can be basically used to build and visualize a better design for 2-parameter ImpLogit model.

The ranges of the covariate x_i are suggested to be lengthened and fixed around the prior believed median value of the logistic distribution. Having larger number of covariate values (m), provided that the sample size of binary observations is fixed, results in retarding the effect of lengthening x_i ranges on the lengths of interval estimates of β_0 and

β_1 . That is, with larger m , the interval estimates are longer. Therefore, short covariate ranges and large m result in more imprecision in interval estimates of β_0 and β_1 . Applying the 2-parameter ImpLogit model, say for a number of dose levels in a dose-response experiment, becomes better by having a small number of dose measured levels.

An important point in fitting 2-parameter ImpLogit model, under IDM, is to recommend ν ranges to shorten the interval estimates. The same thing is for IBLnM imprecise prior where σ_i , $i = 1, 2$, is found to play a stronger role than μ_i in controlling the interval estimates lengths of regression parameters. It is recommended for ν to belong to $[1, 2]$. For 2-parameter ImpLogit model under IBLnM, increasing values of the scale parameters σ_i gives less imprecision in the regression parameters estimates. The preference for each σ_i is to take values greater than 1 to shorten the interval estimates. For the location parameters μ_i , they are not suggested to be far from each other. In bivariate logit-normal distribution, $E \left[\log \left(\frac{\theta_i}{1 - \theta_i} \right) \right] = \mu_i$, and if $\theta_i = 0.5$, there is $\log \left(\frac{\theta_i}{1 - \theta_i} \right) = 0$. Then μ_i 's are suggested to be closed to zero value from both sides.

In 2-parameter ImpLogit model, the lower and upper estimates of β_0 and β_1 do not necessarily correspond to lower and upper estimates of θ . The correspondence occurs for all $x_i \geq 0$, then the covariate x_i is better to have positive values.

The previous mentioned factors are found to have a greater impact on β_1 than β_0 . Changes in lengths of β_1 interval estimates are larger than those for β_0 . This approves the strong effect of the selected covariate ranges on the produced imprecision, since the covariate x_i associates with β_1 in the model structure.

The study of 2-parameter ImpLogit model was enhanced by using the Fisher information matrix (FIM) in Chapter 4. The covariate values that optimize the estimates of the predictive probabilities θ_i , $\forall i$, were computed. An FIM design was compared to the previous ImpLogit designs in Chapter 4. FIM indicated that a covariate range of $x_1 = -3.85$ and $x_2 = 3.85$ optimizes the 2-parameter logistic regression model. This covariate range was applied to minimize interval estimates lengths. The covariate range of $x_1 = -3.85$ and $x_2 = 3.85$ included the range of $x_1 = -2.5$, $x_2 = 2.5$ at which shortest interval estimates were produced for β_0 and β_1 .

Multiple parameter ImpLogit model was applied in Chapter 4 for 3 parameters β_0 , β_1 and β_2 and two covariates x_{i1} and x_{i2} . The behaviour of 3-parameter ImpLogit model

does not differ from the 2-parameter case. Interval estimates of β_0 , β_1 and β_2 become longer with larger m , given that m is the same for both covariates x_{i1} and x_{i2} . Having longer ranges of x_{i1} and x_{i2} produced shorter interval estimates.

5.2 Future Plan

A future plan focuses on multiple parameters ImpLogit model. The question is: how each regression parameter, out of $p + 1$ parameters, affects the produced imprecision? It is important to see how ranges of all covariates can be fixed to have shorter interval estimates of all regression parameters. Therefore, Section 4.10 can be extended to study the relation among all covariates and their effect on the produced imprecision of β_0, \dots, β_p estimates.

Another attractive direction is to model the binomial parameter θ as the cdf of the normal distribution

$$\theta_i = \Phi \left(\mathbf{x}_i' \boldsymbol{\beta} \right), \quad i = 1, \dots, m. \quad (5.1)$$

This model is called the “probit” model. Probit model was first presented by Bliss (1935). For more information about the probit model, McCullagh and Nelder (1989) and Albert and Chib (1993) are recommended.

Probabilistic imprecision can be involved in the probit model to be referred to by ImpProbit model. A comparison between ImpProbit and ImpLogit models allows to look at conditions that control interval estimates lengths of regression parameters. ImpProbit model helps to look at the recommended designs in imprecise logistic regression in parallel with ImpLogit model.

BIBLIOGRAPHY

- [1] Agresti, A. (2007). *An Introduction to Categorical Data Analysis*, 2nd ed. Wiley, New Jersey.
- [2] Aitchison, J., and Shen, S. (1980). Logistic-normal distributions: some properties and uses. *Biometrika*, 67(2): 261 – 272.
- [3] Albert, J., and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422): 669 – 679.
- [4] Aldrich, J. (2008). R. A. Fisher on Bayes and Bayes' Theorem. *Bayesian Analysis*, 3(1): 161 – 170.
- [5] Aughenbaugh, J., and Paredis, C. (2006). The value of using imprecise probabilities in engineering design. *Journal of Mechanical Design*, 128(4): 969 – 979.
- [6] Augustin, T., and Hable, R. (2010). On the impact of robust statistics on imprecise probability models: a review. *Structural Safety*, 32(6): 358 – 365.
- [7] Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, Berlin.
- [8] Berger, J. (1990). Robust Bayesian analysis: sensitivity to the prior. *Journal of Statistical Planning and Inference*, 25: 303 – 328.
- [9] Berger, J. (1994). An overview of robust Bayesian analysis. *Test*, 3(1): 5 – 124.
- [10] Berger, J., and Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, 2(3): 317 – 352

- [11] Berger, J., and Sellke, T. (1987). Testing a point null hypothesis: the irreconcilability of significance levels and evidence. *Journal of the American Statistical Association*, 82(397): 112 – 122.
- [12] Berkson, J. (1944). Application of the logistic function to bio-assay. *Journal of the American Statistical Association*, 39(227): 357 – 365.
- [13] Bernard, J. (2005). An introduction to the imprecise Dirichlet model for Multinomial data. *International Journal of Approximate Reasoning*, 39: 123 – 150.
- [14] Bernardo, J., and Smith A. (1994). *Bayesian Theory*. Wiley, New York.
- [15] Bickis, M. (2009). The imprecise logit-normal model and its application to estimating hazard functions. *Journal of Statistical Theory and Practice*, 3(1): 183 – 195.
- [16] Bickis, M., and Bickis, U. (2007). Predicting the next pandemic: An exercise in imprecise hazards. In De Cooman G, Vejnarova J and Zaffalon M (eds.), *Proceedings of the 5th International Symposium on Imprecise Probability : Theories and Applications (ISIPTA07)*, Prague, Czech Republic, 41 – 46.
- [17] Bliss, C. (1935). The calculation of the dosage-mortality curve. *Annals of Applied Biology*, 22(1): 134 – 167.
- [18] Bloch, D., and Watson, G. (1967). A Bayesian study of the multinomial distribution. *Annals of Mathematical Statistics*, 38(5): 1423 – 1435.
- [19] Box, G., and Tiao, G. (1961). A Further Look at Robustness Via Bayes’s Theorem. *Biometrika*, 49(3-4): 419 – 432.
- [20] Carlin, B., and Sargent, D. (1996). Robust Bayesian approaches for clinical trial monitoring. *Statistics in Medicine*, 15: 1093 – 1106.
- [21] Casella, G., and Berger, R. (2002). *Statistical Inference*, 2nd ed. Duxbury Advanced Series.
- [22] Collet, D. (1991). *Modelling Binary Data*. Chapman and Hall, London.

- [23] Coolen, F. (1993). Imprecise conjugate prior densities for the one-parameter exponential family of distributions. *Statistics and Probability Letters*, 16(5): 337 - 342.
- [24] Coolen, F. (1997). An imprecise Dirichlet model for Bayesian analysis of failure data including right-censored observations. *Reliability Engineering and System Safety*, 56: 61 - 68.
- [25] Cox, D. (1970). *The Analysis of Binary Data*. Methuen, London.
- [26] Cox, D., and Snell, E. (1989). *Analysis of Binary Data*, 2nd ed. Chapman and Hall, London.
- [27] Cozman, F. (2000). Credal networks. *Artificial Intelligence*, 120: 199 – 233.
- [28] Cozman, F. (2005). Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning*, 39: 167 - 184.
- [29] Dale, A. (1999). *A History of Inverse Probability from Thomas Bayes to Karl Pearson*. Springer, New York.
- [30] De Cooman, G. (2002). Precision – imprecision equivalence in a broad class of imprecise hierarchical uncertainty models. *Journal of Statistical Planning and Inference*, 105: 175 – 198.
- [31] De Cooman, G., and Hermans, F. (2008). Imprecise probability trees: bridging two theories of imprecise probability. *Artificial Intelligence*, 172(11): 1400 – 1427.
- [32] De Cooman, G., Hermans, F., and Quaeghebeur, E. (2009). Imprecise Markov chains and their limit behavior. *Probability in the Engineering and Informational Sciences*, 23: 597 – 635.
- [33] de Finetti, B. (1937). La Prévision: Ses Lois Logiques, Ses Sources Subjectives. *Annales de l'Institut Henri Poincaré*, 7: 1 – 68. Translated as Foresight: “Its Logical Laws, Its Subjective Sources,” in Kyburg H and Smokler H. (1964). *Studies in Subjective Probability*. Wiley, New York.

- [34] De Morgan, A. (1837). *Review of Laplace's Théorie Analytique des Probabilités*, 3rd ed. Dublin Review.
- [35] Dempster, A. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38(2): 325 – 339.
- [36] Duda, R., and Hart, P. (1973). *Pattern Classification and Scene Analysis*. Wiley, New York.
- [37] Edgeworth, F. (1883). The method of least squares. *Philosophical Magazine*, 5th series, 34: 190 – 204.
- [38] Edwards, W., Lindman, H., and Savage, L. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70(3): 193 – 242.
- [39] Fienberg, S. (2006). When did Bayesian inference become “Bayesian”? *Bayesian Analysis*, 1(1): 1 – 40.
- [40] Finney, D. (1978). *Statistical Methods in Biological Assay*, 3rd ed. Griffin, London.
- [41] Fisher, R. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London*, Series A, 222: 309 – 368.
- [42] Fisher, R. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.
- [43] Fisher, R. (1956). *Statistical Methods and Scientific Inference*. Oliver and Boyd, Edinburgh.
- [44] Frederic, P., and Lad, F. (2008). Two moments of the logitnormal distribution. *Communications in Statistics-Simulation and Computation*, 37: 1263 – 1269.
- [45] Goldstein, M. (2006). Subjective Bayesian analysis: principles and practice. *Bayesian Analysis*, 1(3): 403 – 420.
- [46] Good, I. (1976). The Bayesian influence, or how to sweep subjectivism under the carpet. *Foundations of Probability Theory, Statistical Inference, and Statistical The-*

- ories of Science*, in Hooker, C. A. and Harper, W. (eds.), Dordrecht, Holland, D. Reidel, 2: 125 – 174.
- [47] Govindarajulu, Z. (1988). *Statistical Techniques in Bioassay*. Karger, Basel.
 - [48] Greenhouse, J., and Wasserman, L. (1995). Robust Bayesian methods for monitoring clinical trials. *Statistics in Medicine*, 14: 1379 – 1391.
 - [49] Gustafson, P. (1996a). Robustness considerations in Bayesian analysis. *Statistical Methods in Medical Research*, 5: 357 – 373.
 - [50] Gustafson, P. (1996b). The effect of mixing distribution misspecification in conjugate mixture models. *Canadian Journal of Statistics*, 24(3): 307 – 318.
 - [51] Hable, R. (2010). Minimum distance estimation in imprecise probability models. *Journal of Statistical Planning and Inference*, 140(2): 461 – 479.
 - [52] Hastings, W. (1970). Monte carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1): 97 – 109.
 - [53] Hilbe, J. (2009). *Logistic Regression Models*. Chapman and Hall/CRC Press.
 - [54] Jeffreys, H. (1939). *Theory of Probability*. Oxford Classic Texts in the Physical Sciences.
 - [55] Keynes, J. (1921). *A Treatise on probability*. Macmillan, London.
 - [56] Kolmogorov, A. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, New York.
 - [57] Kolmogorov, A. (1950). *Foundations of the Theory of Probability*. Chelsea Publishing Company, New York.
 - [58] Kotz, S., Balakrishnan, N., and Johnson, N. (2000). *Continuous Multivariate Distributions. Models and Applications*, Vol. 1. Wiley, New York.
 - [59] Laplace P. (1774). Mémoire sur la probabilit des causes par les événements. *Mémoires de Mathématique et de Physique Présentés a l'Académie Royale des Sciences*, Par Divers Savans, & Lus dans ses Assemblées, 6: 621 – 656.

- [60] Laplace, P. (1812). *Théorie Analytique des Probabilités*. Courcier.
- [61] Lehmann, E., and Casella, G. (1998). *Theory of Point Estimation*, 2nd ed. Springer, New York.
- [62] Leonard, T. (1973). A Bayesian method for histograms. *Biometrika*, 60(2): 297 – 308.
- [63] Lindley, D. (1958). Fiducial distributions and Bayes’ theorem. *Journal of the Royal Statistical Society, Series B*, 20(1): 102 – 107.
- [64] Machina, M., and Schmeidler, D. (1992). A more robust definition of subjective probability. *Econometrica*, 60(4): 745 – 780.
- [65] Troffaes, M. (2007). Decision making under uncertainty using imprecise probabilities. *International Journal of Approximate Reasoning*, 45(1): 17 – 29.
- [66] McCullagh, C., and Nelder, J. (1989). *Generalized Linear Models*. Chapman and Hall, London.
- [67] McCullagh, C., and Searle, S. (2001). *Generalized, Linear, and Mixed Models*. Wiley, New York.
- [68] Mead, R. (1965). A generalized logit-normal distribution. *Biometrics*, 21(3): 721 – 732.
- [69] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6): 1087 - 1092.
- [70] Ostrogradsky, M. (1838). Extrait d’un mémoire sur la probabilité des erreurs des tribunaux. *Mémoires de l’Académie impériale des Sciences de St. Petersburg*, Ser. 6, 3, xix-xxv.
- [71] Patterson, K. (1987). *Pandemic Influenza, 1700 - 1900 : A Study in Historical Epidemiology*. Rowman and Littlefield, Totowa, NJ.

- [72] Pearson, K. (1907). On the influence of past experience on future expectation. *Philosophical Magazine*, 13(6): 365 – 378.
- [73] Pericchi, L., and Walley, P. (1991). Robust Bayesian credible intervals and prior ignorance. *International Statistical Review*, 58(1): 1 – 23.
- [74] Quaeghebeur, E., and De Cooman, G. (2005). Imprecise probability models for inference in exponential families. In Cozman, F., Nau, R., and Seidenfeld, T., (eds.), *Proceedings of the 4th International Symposium on Imprecise Probabilities and their Applications (ISIPTA05)*, Pittsburgh, Pennsylvania.
- [75] Ramsey, F. (1926). *Truth and Probability*. Published in Braithwaite, R. (1931). *Foundations of Mathematics and Other Logical Essays*. Kegan, Paul, Trench, Trubner & Co., London, Ch. VII, 156 – 198. Reprinted in Kyburg, H., and Smokler, H. (1964). *Studies in Subjective Probability*. Wiley, New York, 61 – 92.
- [76] Robert, C. (2001). *The Bayesian Choice : from Decision-Theoretic Foundations to Computational Implementation*, 2nd ed. Springer, New York.
- [77] Robert, C., and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer, New York.
- [78] Savage, L. (1954). *The Foundations of Statistics*. Wiley, New York.
- [79] Savage, L. (1961). The foundations of statistics reconsidered. In Neyman, J., ed, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1: 575 – 586.
- [80] Savage, L. (1962). *The Foundations of Statistical Inference. A Discussion*. Barnard, G., and Cox, D., eds, Methuen, London.
- [81] Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press.
- [82] Shafer, G., and Vovk, V. (2001). *Probability and Finance: It's Only a Game!*. Wiley, New York.

- [83] Spiegelhalter, D., Freedman, L., and Parmar, M. (1994). *Bayesian approaches to randomized trials*. *Journal of the Royal Statistical Society*, Series A, 157(3): 357 – 416.
- [84] Stigler, S. (1978). Francis Ysidro Edgeworth, Statistician. *Journal of the Royal Statistical Society*, Series A, 141(3): 287 – 322.
- [85] Stigler, S. (1983). Who discovered Bayes’s theorem? *The American Statistician*, 37(4): 290 – 296.
- [86] Stigler, S. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge, MA, and London: Belkap Press of Harvard University Press.
- [87] Tiao, G., and Tan, W. (1965). Bayesian analysis of random-effect models in the analysis of variance. I. Posterior distribution of variance-components. *Biometrika*, 52(1 – 2): 37 – 53.
- [88] Vicig, P. (2008). Imprecise probabilities in finance and economics. *International Journal of Approximate Reasoning*, 49(1): 99 – 100.
- [89] Walley, P. (1991). *Statistical Reasoning With Imprecise Probabilities*. Chapman and Hall, London, 42.
- [90] Walley, P. (1996). Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society*, Series B, 58(1): 3 – 57.
- [91] Walley, P., Gurrin, L., and Burton, P. (1996). Analysis of clinical data using imprecise prior probabilities. *The Statistician*, 45(4): 457 – 485.
- [92] Walter, G., Augustine, T., and Peters, A. (2007). Linear regression analysis under sets of conjugate priors. In De Cooman G, Vejnarova J and Zaffalon M (eds.), *Proceedings of the 5th International Symposium on Imprecise Probability: Theories and Applications (ISIPTA07)*, Prague, Czech Republic.
- [93] Wasserman, L. (1992). Recent methodological advances in robust Bayesian inference. In Bernardo, J., Berger, J., Dawid, A., and Smith, A., eds, *Bayesian Statistics 4*, Oxford, Oxford University Press, 483 – 502.

- [94] Weichselberger, K. (2000). The theory of interval probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24(2 – 3): 149 – 170.
- [95] Zadeh, L. (1975). The concept of a linguistic variable and its applications to approximate reasoning-I. *Information Sciences*, 8: 199 – 249.
- [96] Zadeh, L. (2002). Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. *Journal of Statistical Planning and Inference*, 105(1): 233 – 264.
- [97] Zadeh, L. (2006). Generalized theory of uncertainty (GTU): principal concepts and ideas. *Computational Statistics and Data Analysis*, 51(1): 15 – 46.
- [98] Zaffalon, M. (2002). The naive credal classifier. *Journal of Statistical Planning and Inference*, 105(1): 5 – 21.

APPENDIX A

R CODE

```
#####  
## This program is for MCMC method to estimate regression ##  
## is logit model. #####  
require(xtable);library(MCMCpack);library(MASS);  
library(coda);library(lattice) ## some required packages.  
##### required info #####  
par(mfrow=c(3,3),mar=c(5,4.3,.31,1),oma=c(2,2,1,1))  
nu = c(0.5,1,2)      ### concentration parameter value in IDM.  
df = 5               ### degrees of freedom of t-distribution.  
numdraws = 2000      ### number of MCMC draws.  
firstdose = 0; seconddose = 10.2 ### first and second dose  
numinterval = 5      ### m = 6 dose levels  
##### the jacobian of the prior induced pdf #####  
prior5 <- function (b0,b1,xf,xs) {  
  z = abs(xs-xf)*(exp(b0+xf*b1)/((1+exp(b0+xf*b1))*2))  
  w = (exp(b0+xs*b1)/((1+exp(b0+xs*b1))*2))  
  z*w}  
#####functions end#####  
x = c(0, 2.6, 3.8, 5.1, 7.7, 10.2)      ### amounts of doses.  
n = c(49, 50, 48, 46, 49, 50)      ### number of experimented insects.  
yobs = c(0, 6, 16, 24, 42, 44)      ### number of affected insects.  
m = length(n)      ### m value.  
##### to solve the data in classical approach #####  
probobs = yobs/n
```

```

d = data.frame(cbind(probobs,x))
logit.fit = glm(probobs~x, data = d, family=binomial("logit"))
logit.fit$coefficients[1]
logit.fit$coefficients[2]
#####
q = c(0.15,.3,.45);
f = rep(0,length(q)); s = rep(0,length(q))
b0hat = matrix(rep(0,length(nu)*length(q)),length(nu), length(q))
b1hat = matrix(rep(0,length(nu)*length(q)),length(nu), length(q))
corr = matrix(rep(0,length(nu)*length(q)),length(nu), length(q))
sigma1 = 0.2;sigma2 = c(0.3,0.3,0.3)
for (h in 1:length(nu)){
nu1 = rep(0,length(q)); nu2 = rep(0,length(q));
for (k in 1:length(q)){
nu1[k]=nu[h]*q[k]      ;      nu2[k]=nu[h]*(1-q[k])
f[k] = round(q[k]*m)      ###   the sequential order for theta j
s[k] = round((1.1-q[k])*m) ###   the sequential order for theta k
b0cand = numeric(numdraws)
b1cand = numeric(numdraws)
randomt1 = numeric(numdraws)
randomt2 = numeric(numdraws)
jointcand = numeric(numdraws)
randomt1[1] = rt(1,df)
randomt2[1] = rt(1,df)
b0cand[1] = randomt1[1]*sigma1+logit.fit$coefficients[1]
b1cand[1] = randomt2[1]*sigma2[h]+logit.fit$coefficients[2]
b0 = numeric(numdraws)
b1 = numeric(numdraws)
b0[1] = b0cand[1]

```



```

b1[1] = b1cand[1]
##### main body of the code#####
##### to calculate the joint value for the initial point ###
likecand = numeric(length(x))
priorcand = numeric(numdraws)
likelihood = numeric(numdraws)
comptheta1 = numeric(length(x));comptheta2 = numeric(length(x))
comptheta = numeric(length(x))
for (i in 1:m){
  comptheta1[i] = exp(b0cand[1]+x[i]*b1cand[1])
  comptheta2[i] = (1+exp(b0cand[1]+x[i]*b1cand[1]))
  comptheta[i] = comptheta1[i]/comptheta2[i]
  likecand[i] = dbinom(yobs[i],n[i],comptheta[i])
}
d1 = exp(b0cand[1]+x[f[k]]*b1cand[1])
d2 = (1+exp(b0cand[1]+x[f[k]]*b1cand[1]))
d3 = exp(b0cand[1]+x[s[k]]*b1cand[1])
d4 = (1+exp(b0cand[1]+x[s[k]]*b1cand[1]))
e1 = d1/d2
e2 = d3/d4
e3 = e2-e1
e4 = 1 - e2
ee = c(e1,e3,e4)
density = ddirichlet(ee,c(nu1[k],nu2[k]-nu1[k],nu[h]-nu2[k]))
priorcand[1] = density*(prior5(b0cand[1],b1cand[1],x[f[k]],x[s[k]]))
jointcand[1] = prod(likecand)*priorcand[1]
#####the rest of the chain #####
alphamcmc1 = numeric(numdraws)
alphamcmc2 = numeric(numdraws)

```

```

alphamcmc = numeric(numdraws)
for (j in 2:numdraws){
  randomt1[j] = rt(1,df)
  b0cand[j] = randomt1[j]*sigma1+logit.fit$coefficients[1]
  randomt2[j] = rt(1,df)
  b1cand[j] = randomt2[j]*sigma2[h]+logit.fit$coefficients[2]
  for (i in 1:m){
    comptheta1[i] = exp(b0cand[j]+x[i]*b1cand[j])
    comptheta2[i] = (1+exp(b0cand[j]+x[i]*b1cand[j]))
    comptheta[i] = comptheta1[i]/comptheta2[i]
    likecand[i] = dbinom(yobs[i],n[i],comptheta[i])
  }
  d1 = exp(b0cand[j]+x[f[k]]*b1cand[j])
  d2 = (1+exp(b0cand[j]+x[f[k]]*b1cand[j]))
  d3 = exp(b0cand[j]+x[s[k]]*b1cand[j])
  d4 = (1+exp(b0cand[j]+x[s[k]]*b1cand[j]))
  e1 = d1/d2
  e2 = d3/d4
  e3 = e2-e1
  e4 = 1 - e2
  ee = c(e1,e3,e4)
  density = ddirichlet(ee,c(nu1[k],nu2[k]-nu1[k],nu[h]-nu2[k]))
  priorcand[j]=density*(prior5(b0cand[j],b1cand[j],x[f[k]],x[s[k]]))
  likelihood[j] = prod(likecand)
  jointcand[j] = likelihood[j]*priorcand[j]
  alphamcmc1[j]=jointcand[j]
  alphamcmc2[j]=jointcand[j-1]
  ### to find alpha value in Metropolis Hastings.
  alphamcmc[j]=alphamcmc1[j]/alphamcmc2[j]
}

```

```

if (runif(1,0,1) < min(alphamcmc[j],1)){
b0cand[j] = b0cand[j]
b1cand[j] = b1cand[j]
jointcand[j] = jointcand[j]
}
else{
b0cand[j] = b0cand[j-1]
b1cand[j] = b1cand[j-1]
jointcand[j] = jointcand[j-1]
}
b0[j] = b0cand[j]
b1[j] = b1cand[j]
likecand = numeric(length(x))
comptheta = numeric(length(x))
}

##### Metropolis Hastings ends #####

b0hat[h,k] = mean(b0)
b1hat[h,k] = mean(b1)

##### the plots #####
x1 = c(seq(min(x)-2,max(x)+2,0.1))
P = matrix(rep(0,length(x1)))
plot(x,yobs/n,pch=20,xlim=c(-2,13),ylim=c(0,1),las=1,ylab="")
x1 = c(seq(min(x)-2,max(x)+2,0.05))
P = matrix(rep(0,length(x1)))
thetaestimate1 = rep(0,length(x1));
thetaestimate2 = rep(0,length(x1))
thetaestimate = rep(0,length(x1))
for (i in 1:length(x1)){
thetaestimate1[i] = (exp(b0hat[h,k]+x1[i]*b1hat[h,k]))

```

```

thetaestimate2[i] = (1+exp(b0hat[h,k]+x1[i]*b1hat[h,k]))
thetaestimate[i] = thetaestimate1[i]/thetaestimate2[i]
}

lines(x1,thetaestimate,pch=20,lty=1,col=2,xlim=c(-2,13),ylim=c(0,1))
text(-1,0.9,expression(hat(beta)[0]))
text(.5,0.9,"=")
text(3.5,0.9,round(b0hat[h,k],3))
text(-1,0.7,expression(hat(beta)[1]))
text(.5,0.7,"=")
text(3.3,0.7,round(b1hat[h,k],3))
mtext(expression(nu),at=-2,3,cex=1.5)
mtext("=",at=0,3,lwd=1.5)
mtext(nu[h],at=3,3,lwd=1.5)
mtext(",",at=5,3,cex=1.5)
mtext("Pair:",at=8,3,lwd=1.5,cex=1)
mtext(f[k],at=11,3,lwd=1.5)
mtext(",",at=12,3,lwd=1.5)
mtext(s[k],at=13,3,lwd=1.5)
#####
}
}

#####code is finished#####

table1 = xtable(cbind(x,n,yobs), digits=3)
colnames(table1) <- c("$Dose$","$n$", "$y_{obs}$")
print(table1, sanitize.text.function = function(x) { x })

table2 = xtable(b0hat, digits=3)
c1 = c("$\\theta_{1}$, $\\theta_{6}$\\left(x_{1},x_{6}\\right)$")
c2 = c("$\\theta_{2}$, $\\theta_{5}$\\left(x_{2},x_{5}\\right)$")
c3 = c("$\\theta_{3}$, $\\theta_{4}$\\left(x_{3},x_{4}\\right)$")

```

```

rownames(table2) <- c(c1,c2,c3)
colnames(table2) <- c("$\\nu=0.5$", "$\\nu=1$", "$\\nu=2$")
table2 = xtable(table2, digits=3)
align(table2) = "cccc"
print(table2, sanitize.text.function = function(x) { x })
table3 = xtable(b1hat, digits=3)
c1 = c("$\\theta_{1}$, $\\theta_{6}$\\left(x_{1},x_{6}\\right)$")
c2 = c("$\\theta_{2}$, $\\theta_{5}$\\left(x_{2},x_{5}\\right)$")
c3 = c("$\\theta_{3}$, $\\theta_{4}$\\left(x_{3},x_{4}\\right)$")
rownames(table3) <- c(c1,c2,c3)
colnames(table3) <- c("$\\nu=0.5$", "$\\nu=1$", "$\\nu=2$")
table3 = xtable(table3, digits=3)
align(table3) = "cccc"
print(table3, sanitize.text.function = function(x) { x })

```

APPENDIX B

POSTERIOR LOGIT-NORMAL DENSITY

The logit-normal distribution is related to the normal distribution as follows

$$\text{if } \theta \sim \text{logit-normal}(\mu, \sigma^2), \quad (\text{B.1})$$

and

$$\text{if } \theta = \frac{e^w}{1 + e^w}, \quad (\text{B.2})$$

then

$$w = \log\left(\frac{\theta}{1 - \theta}\right) \sim N(\mu, \sigma^2). \quad (\text{B.3})$$

To find the distribution of w conditional on y , where y comes from a binomial random variable $\text{bin}(n, \theta)$, the posterior pdf is derived as follows

$$\begin{aligned} \pi(w|y) &= \frac{f(y|w) \cdot \pi(w)}{m(y)} = \frac{\text{binomial} \times \text{normal}}{\text{marginal}} \\ &= \frac{\left[\binom{n}{y} \left(\frac{e^w}{1 + e^w} \right)^y \left(1 - \frac{e^w}{1 + e^w} \right)^{n-y} \right] \cdot \left\{ \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp \left[-\frac{(w - \mu)^2}{2\sigma^2} \right] \right\}}{m(y)} \\ &\propto \left[\left(\frac{e^w}{1 + e^w} \right)^y \left(1 - \frac{e^w}{1 + e^w} \right)^{n-y} \right] \cdot \left[\exp \left(-\frac{(w - \mu)^2}{2\sigma^2} \right) \right] \\ &= \left[\frac{e^{wy}}{(1 + e^w)^n} \right] \left\{ \exp \left[-\frac{(w - \mu)^2}{2\sigma^2} \right] \right\} = \frac{\exp \left[\frac{-(w - \mu)^2 + 2\sigma^2 wy}{2\sigma^2} \right]}{(1 + e^w)^n} \\ &= \frac{\exp \left[\frac{-(w^2 + \mu^2 - 2w\mu) + 2\sigma^2 wy}{2\sigma^2} \right]}{(1 + e^w)^n} \end{aligned}$$

$$\begin{aligned}
&= \frac{\exp\left(\frac{-w^2 - \mu^2 + 2w\mu + 2\sigma^2 wy}{2\sigma^2}\right)}{(1 + e^w)^n} \\
&\propto \frac{\exp\left[\frac{-w^2 - \mu^2 + 2w\mu + 2\sigma^2 wy - 2\mu\sigma^2 y - (\sigma^2 y)^2}{2\sigma^2}\right]}{(1 + e^w)^n} \\
&= \frac{\exp\left\{\frac{-[w - (\mu + \sigma^2 y)]^2}{2\sigma^2}\right\}}{(1 + e^w)^n}. \tag{B.4}
\end{aligned}$$

Now, the posterior distribution of θ is found by writing (B.4) in terms of θ , where based on (B.3), there is

$$\begin{aligned}
\pi(\theta|y) &\propto \frac{\exp\left\{\frac{-\left[\log\left(\frac{\theta}{1-\theta}\right) - (\mu + \sigma^2 y)\right]^2}{2\sigma^2}\right\}}{\left\{1 + \exp\left[\log\left(\frac{\theta}{1-\theta}\right)\right]\right\}^n} \cdot \left|\frac{dw}{d\theta}\right| \\
&\propto \frac{\exp\left\{\frac{-\left[\log\left(\frac{\theta}{1-\theta}\right) - (\mu + \sigma^2 y)\right]^2}{2\sigma^2}\right\}}{\left\{1 + \exp\left[\log\left(\frac{\theta}{1-\theta}\right)\right]\right\}^n} \cdot \frac{1}{\theta(1-\theta)} \\
&\propto \frac{\exp\left\{\frac{-\left[\log\left(\frac{\theta}{1-\theta}\right) - (\mu + \sigma^2 y)\right]^2}{2\sigma^2}\right\}}{\left(\frac{1}{1-\theta}\right)^n} \cdot \frac{1}{\theta(1-\theta)} \\
&\propto \exp\left\{\frac{-\left[\log\left(\frac{\theta}{1-\theta}\right) - (\mu + \sigma^2 y)\right]^2}{2\sigma^2}\right\} \cdot \frac{(1-\theta)^{n-1}}{\theta}. \tag{B.5}
\end{aligned}$$

Both of the prior and posterior distributions in (2.40) and (B.5) belong to exponential families. This can be shown by taking the log of the prior density in (2.40) so that it is

is written as

$$\begin{aligned}
& \log [f(\theta|\mu, \sigma)] \\
&= \log \left\{ \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp \left\{ -\frac{\left[\log\left(\frac{\theta}{1-\theta}\right) - \mu \right]^2}{2\sigma^2} \right\} \cdot \left[\frac{1}{\theta(1-\theta)} \right] \right\} \\
&= \log \left(\frac{1}{\sigma\sqrt{2\pi}} \right) - \frac{\left[\log\left(\frac{\theta}{1-\theta}\right) - \mu \right]^2}{2\sigma^2} + \log \left[\frac{1}{\theta(1-\theta)} \right] \\
&= \log \left(\frac{1}{\sigma\sqrt{2\pi}} \right) - \frac{\left[\log\left(\frac{\theta}{1-\theta}\right) \right]^2 + \mu^2 - 2\mu \log\left(\frac{\theta}{1-\theta}\right)}{2\sigma^2} \\
&+ \log \left[\frac{1}{\theta(1-\theta)} \right] \\
&= \log \left(\frac{1}{\sigma\sqrt{2\pi}} \right) - \frac{1}{2\sigma^2} \left[\log\left(\frac{\theta}{1-\theta}\right) \right]^2 - \frac{\mu^2}{2\sigma^2} + \frac{2\mu}{2\sigma^2} \left[\log\left(\frac{\theta}{1-\theta}\right) \right] \\
&+ \log \left[\frac{1}{\theta(1-\theta)} \right] \\
&= \log \left(\frac{1}{\sigma\sqrt{2\pi}} \right) - \frac{1}{2\sigma^2} \left[\log\left(\frac{\theta}{1-\theta}\right) \right]^2 - \frac{\mu^2}{2\sigma^2} \\
&+ \frac{2\mu}{2\sigma^2} [\log(\theta) - \log(1-\theta)] - \log(\theta) - \log(1-\theta) \\
&= \left[\underbrace{\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{\mu^2}{2\sigma^2}}_{\eta_0} \right] + \left\{ \underbrace{-\frac{1}{2\sigma^2}}_{\eta_1} \underbrace{\left[\log\left(\frac{\theta}{1-\theta}\right) \right]^2}_{\tau_1} \right\} \\
&+ \left[\underbrace{\left(\frac{\mu}{\sigma^2} - 1 \right)}_{\eta_2} \underbrace{\log(\theta)}_{\tau_2} \right] + \left[\underbrace{-\left(\frac{\mu}{\sigma^2} + 1 \right)}_{\eta_3} \underbrace{\log(1-\theta)}_{\tau_3} \right] \\
&= \eta_0(\mu, \sigma) + \eta_1(\mu, \sigma) \tau_1(\theta) + \eta_2(\mu, \sigma) \tau_2(\theta) + \eta_3(\mu, \sigma) \tau_3(\theta). \tag{B.6}
\end{aligned}$$

It is clear that (B.6) forms an exponential family since η_0 , η_1 , η_2 and η_3 are the

hyperparameters and τ_1 , τ_2 and τ_3 are the natural functions of θ where

$$\begin{cases} \eta_1 &= -\frac{1}{2\sigma^2}, \\ \eta_2 &= \frac{\mu}{\sigma^2} - 1, \\ \eta_3 &= -\frac{\mu}{\sigma^2} - 1. \end{cases} \quad (\text{B.7})$$

The log of the posterior pdf given in (B.5) is also taken and compared to the log of the prior density in (B.6). That is

$$\begin{aligned} \log [\pi (\theta|y)] &= \log \left\{ C \cdot \exp \left\{ \frac{- \left[\log \left(\frac{\theta}{1-\theta} \right) - (\mu + \sigma^2 y) \right]^2}{2\sigma^2} \right\} \cdot \frac{(1-\theta)^{n-1}}{\theta} \right\} \\ &= K + \frac{- \left[\log \left(\frac{\theta}{1-\theta} \right) - (\mu + \sigma^2 y) \right]^2}{2\sigma^2} + \log \left[\frac{(1-\theta)^{n-1}}{\theta} \right] \\ &= K - \frac{1}{2\sigma^2} \left\{ \left[\log \left(\frac{\theta}{1-\theta} \right) \right]^2 + (\mu + \sigma^2 y)^2 - 2 (\mu + \sigma^2 y) \left[\log \left(\frac{\theta}{1-\theta} \right) \right] \right\} \\ &\quad + \log \left[\frac{(1-\theta)^{n-1}}{\theta} \right] \\ &= \left[K - \frac{(\mu + \sigma^2 y)^2}{2\sigma^2} \right] - \frac{1}{2\sigma^2} \left[\log \left(\frac{\theta}{1-\theta} \right) \right]^2 + \frac{\mu + \sigma^2 y}{\sigma^2} \left[\log \left(\frac{\theta}{1-\theta} \right) \right] \\ &\quad + (n-1) \log (1-\theta) - \log (\theta) \\ &= \left[K - \frac{(\mu + \sigma^2 y)^2}{2\sigma^2} \right] - \frac{1}{2\sigma^2} \left[\log \left(\frac{\theta}{1-\theta} \right) \right]^2 + \frac{\mu + \sigma^2 y}{\sigma^2} \log (\theta) \\ &\quad - \frac{\mu + \sigma^2 y}{\sigma^2} \log (1-\theta) \\ &\quad + (n-1) \log (1-\theta) - \log (\theta) \\ &= \left[K - \frac{(\mu + \sigma^2 y)^2}{2\sigma^2} \right] - \frac{1}{2\sigma^2} \left[\log \left(\frac{\theta}{1-\theta} \right) \right]^2 + \left(\frac{\mu + \sigma^2 y}{\sigma^2} - 1 \right) \log (\theta) \\ &\quad + \left(-\frac{\mu + \sigma^2 y}{\sigma^2} + n - 1 \right) \log (1-\theta) \end{aligned}$$

$$\begin{aligned}
&= \left[\underbrace{K - \frac{(\mu + \sigma^2 y)^2}{2\sigma^2}}_{\eta_0^*} \right] + \left\{ \underbrace{-\frac{1}{2\sigma^2}}_{\eta_1^*} \underbrace{\left[\log \left(\frac{\theta}{1-\theta} \right) \right]^2}_{\tau_1} \right\} \\
&+ \left[\underbrace{\left(\frac{\mu}{\sigma^2} - 1 + y \right)}_{\eta_2^*} \underbrace{\log(\theta)}_{\tau_2} \right] + \left[\underbrace{\left(-\frac{\mu}{\sigma^2} - 1 + n - y \right)}_{\eta_3^*} \underbrace{\log(1-\theta)}_{\tau_3} \right] \\
&= \eta_0^*(\mu, \sigma) + \eta_1^*(\mu, \sigma) \tau_1(\theta) + \eta_2^*(\mu, \sigma) \tau_2(\theta) + \eta_3^*(\mu, \sigma) \tau_3(\theta). \tag{B.8}
\end{aligned}$$

Comparing both of (B.6) and (B.8) shows that

$$\begin{cases} \eta_1^* &= \eta_1, \\ \eta_2^* &= \eta_2 + y, \\ \eta_3^* &= \eta_3 + n - y. \end{cases} \tag{B.9}$$

APPENDIX C

POSTERIOR MULTIVARIATE LOGIT-NORMAL DENSITY

Taking the log of the posterior multivariate logit-normal density gives

$$\begin{aligned}
\log [\pi(\boldsymbol{\theta}|\mathbf{y})] &= K - \frac{1}{2} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) - (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y}) \right]' \boldsymbol{\Sigma}^{-1} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) - (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y}) \right] \\
&\quad + \mathbf{1}' \log \left[\frac{(\mathbf{1} - \boldsymbol{\theta})^{\mathbf{n}-\mathbf{1}}}{\boldsymbol{\theta}} \right] \\
&= K - \frac{1}{2} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) \right]' \boldsymbol{\Sigma}^{-1} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) \right] + \frac{1}{2} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) \right]' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y}) \\
&\quad + \frac{1}{2} (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y})' \boldsymbol{\Sigma}^{-1} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) \right] - \frac{1}{2} (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y}) \\
&\quad + \mathbf{1}' \log [(\mathbf{1} - \boldsymbol{\theta})^{\mathbf{n}-\mathbf{1}}] - \mathbf{1}' \log (\boldsymbol{\theta}) \\
&= K - \frac{1}{2} (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y}) - \frac{1}{2\sigma^2} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) \right]' \mathbf{R}^{-1} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) \right] \\
&\quad + \frac{1}{2\sigma^2} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) \right]' \mathbf{R}^{-1} (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y}) + \frac{1}{2\sigma^2} (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y})' \mathbf{R}^{-1} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) \right] \\
&\quad + \mathbf{1}' \log [(\mathbf{1} - \boldsymbol{\theta})^{\mathbf{n}-\mathbf{1}}] - \mathbf{1}' \log (\boldsymbol{\theta}) \\
&= K - \frac{1}{2\sigma^2} (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y})' \mathbf{R}^{-1} (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y}) - \frac{1}{2\sigma^2} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) \right]' \mathbf{R}^{-1} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) \right] \\
&\quad + \frac{1}{2\sigma^2} [\log (\boldsymbol{\theta}) - \log (\mathbf{1} - \boldsymbol{\theta})]' \mathbf{R}^{-1} (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y}) + \frac{1}{2\sigma^2} (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y})' \mathbf{R}^{-1} [\log (\boldsymbol{\theta}) - \log (\mathbf{1} - \boldsymbol{\theta})] \\
&\quad + (\mathbf{n} - \mathbf{1})' \log [(\mathbf{1} - \boldsymbol{\theta})] - \mathbf{1}' \log (\boldsymbol{\theta}) \\
&= K - \frac{1}{2\sigma^2} (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y})' \mathbf{R}^{-1} (\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{y}) - \frac{1}{2\sigma^2} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) \right]' \mathbf{R}^{-1} \left[\log \left(\frac{\boldsymbol{\theta}}{\mathbf{1} - \boldsymbol{\theta}} \right) \right] \\
&\quad + \frac{1}{2} [\log (\boldsymbol{\theta}) - \log (\mathbf{1} - \boldsymbol{\theta})]' \mathbf{R}^{-1} \left(\frac{1}{\sigma^2} \boldsymbol{\mu} + \mathbf{R} \mathbf{y} \right) + \frac{1}{2} \left(\frac{1}{\sigma^2} \boldsymbol{\mu} + \mathbf{R} \mathbf{y} \right)' \mathbf{R}^{-1} [\log (\boldsymbol{\theta}) - \log (\mathbf{1} - \boldsymbol{\theta})] \\
&\quad + (\mathbf{n} - \mathbf{1})' \log [(\mathbf{1} - \boldsymbol{\theta})] - \mathbf{1}' \log (\boldsymbol{\theta})
\end{aligned}$$

$$\begin{aligned}
&= K - \frac{1}{2\sigma^2} (\boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{y})' \mathbf{R}^{-1} (\boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{y}) - \frac{1}{2\sigma^2} \left[\log\left(\frac{\boldsymbol{\theta}}{1-\boldsymbol{\theta}}\right) \right]' \mathbf{R}^{-1} \left[\log\left(\frac{\boldsymbol{\theta}}{1-\boldsymbol{\theta}}\right) \right] \\
&+ \left[\left(\frac{1}{\sigma^2} \boldsymbol{\mu} + \mathbf{R}\mathbf{y} \right)' \mathbf{R}^{-1} - \mathbf{1}' \right] \log(\boldsymbol{\theta}) - \left[\left(\frac{1}{\sigma^2} \boldsymbol{\mu} + \mathbf{R}\mathbf{y} \right)' \mathbf{R}^{-1} - (\mathbf{n} - \mathbf{1})' \right] \log(1 - \boldsymbol{\theta}) \\
&= \left[\underbrace{K - \frac{1}{2\sigma^2} (\boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{y})' \mathbf{R}^{-1} (\boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{y})}_{\boldsymbol{\eta}_0^*} \right] + \left\{ \underbrace{-\frac{1}{2\sigma^2} \left[\log\left(\frac{\boldsymbol{\theta}}{1-\boldsymbol{\theta}}\right) \right]}' \mathbf{R}^{-1} \left[\log\left(\frac{\boldsymbol{\theta}}{1-\boldsymbol{\theta}}\right) \right]}_{\boldsymbol{\eta}_1^*} \right\} \\
&+ \left[\underbrace{\left(\frac{1}{\sigma^2} \boldsymbol{\mu}' \mathbf{R}^{-1} + \mathbf{y}' - \mathbf{1}' \right)}_{\boldsymbol{\eta}_2^*} \underbrace{\log(\boldsymbol{\theta})}_{\boldsymbol{\tau}_2} \right] + \left\{ \underbrace{-\left[\frac{1}{\sigma^2} \boldsymbol{\mu}' \mathbf{R}^{-1} + \mathbf{y}' - (\mathbf{n} - \mathbf{1})' \right]}_{\boldsymbol{\eta}_3^*} \underbrace{\log(1 - \boldsymbol{\theta})}_{\boldsymbol{\tau}_3} \right\} \\
&= \boldsymbol{\eta}_0^*(\boldsymbol{\mu}, \sigma, \mathbf{R}) + \boldsymbol{\eta}_1^*(\boldsymbol{\mu}, \sigma, \mathbf{R}) \boldsymbol{\tau}_1(\boldsymbol{\theta}) + \boldsymbol{\eta}_2^*(\boldsymbol{\mu}, \sigma, \mathbf{R}) \boldsymbol{\tau}_2(\boldsymbol{\theta}) + \boldsymbol{\eta}_3^*(\boldsymbol{\mu}, \sigma, \mathbf{R}) \boldsymbol{\tau}_3(\boldsymbol{\theta}). \quad (\text{C.1})
\end{aligned}$$

From (C.1), the hyperparameters are

$$\begin{cases} \eta_1^* &= -\frac{1}{2\sigma^2}, \\ \eta_2^* &= \frac{1}{\sigma^2} \boldsymbol{\mu}' \mathbf{R}^{-1} + \mathbf{y}' - \mathbf{1}', \\ \eta_3^* &= -\frac{1}{\sigma^2} \boldsymbol{\mu}' \mathbf{R}^{-1} + \mathbf{n}' - \mathbf{y}' - \mathbf{1}'. \end{cases} \quad (\text{C.2})$$

APPENDIX D

CALCULATIONS AND GRAPHS FROM CHAPTER 3

Table D.1: Interval estimates of β_0 and β_1 that correspond to top plots in Figure D.1 for 2-parameter mixed ImpLogit model with $w = 0.5$.

selected pair	$\hat{\beta}_0$	$\bar{\hat{\beta}}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\hat{\beta}}_1$	$\text{imp}(\hat{\beta}_1)$
$\theta_1, \theta_6(x_1, x_6)$	-3.231	-3.216	0.015	0.584	0.618	0.034
$\theta_2, \theta_5(x_2, x_5)$	-3.232	-3.217	0.015	0.580	0.619	0.039
$\theta_3, \theta_4(x_3, x_4)$	-3.233	-3.218	0.015	0.578	0.619	0.041

Table D.2: Interval estimates of β_0 and β_1 that correspond to middle plots in Figure D.1 for 2-parameter mixed ImpLogit model with $w = 0.8$.

selected pair	$\hat{\beta}_0$	$\bar{\hat{\beta}}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\hat{\beta}}_1$	$\text{imp}(\hat{\beta}_1)$
$\theta_1, \theta_6(x_1, x_6)$	-3.224	-3.222	0.002	0.596	0.619	0.023
$\theta_2, \theta_5(x_2, x_5)$	-3.225	-3.219	0.006	0.594	0.621	0.027
$\theta_3, \theta_4(x_3, x_4)$	-3.227	-3.218	0.009	0.592	0.622	0.030

Table D.3: Interval estimates of β_0 and β_1 that correspond to bottom plots in Figure D.1 for 2-parameter mixed ImpLogit model with $w = 0.9$.

selected pair	$\hat{\beta}_0$	$\bar{\hat{\beta}}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\hat{\beta}}_1$	$\text{imp}(\hat{\beta}_1)$
$\theta_1, \theta_6(x_1, x_6)$	-3.227	-3.220	0.007	0.601	0.615	0.014
$\theta_2, \theta_5(x_2, x_5)$	-3.226	-3.221	0.005	0.598	0.618	0.020
$\theta_3, \theta_4(x_3, x_4)$	-3.227	-3.219	0.008	0.597	0.619	0.022

Table D.4: Interval estimates of θ for Table D.1 when $x = 5$.

selected pair	$\hat{\theta}$	$\bar{\theta}$	$\text{imp}(\hat{\theta})$
$\theta_1, \theta_6(x_1, x_6)$	0.423	0.469	0.046
$\theta_2, \theta_5(x_2, x_5)$	0.418	0.470	0.052
$\theta_3, \theta_4(x_3, x_4)$	0.415	0.469	0.054

Table D.5: Interval estimates of θ for Table D.2 when $x = 5$.

selected pair	$\hat{\theta}$	$\bar{\theta}$	$\text{imp}(\hat{\theta})$
$\theta_1, \theta_6(x_1, x_6)$	0.439	0.468	0.029
$\theta_2, \theta_5(x_2, x_5)$	0.437	0.472	0.035
$\theta_3, \theta_4(x_3, x_4)$	0.434	0.473	0.039

Table D.6: Interval estimates of θ for Table D.3 when $x = 5$.

selected pair	$\hat{\theta}$	$\bar{\theta}$	$\text{imp}(\hat{\theta})$
$\theta_1, \theta_6(x_1, x_6)$	0.445	0.464	0.019
$\theta_2, \theta_5(x_2, x_5)$	0.441	0.468	0.027
$\theta_3, \theta_4(x_3, x_4)$	0.440	0.470	0.030

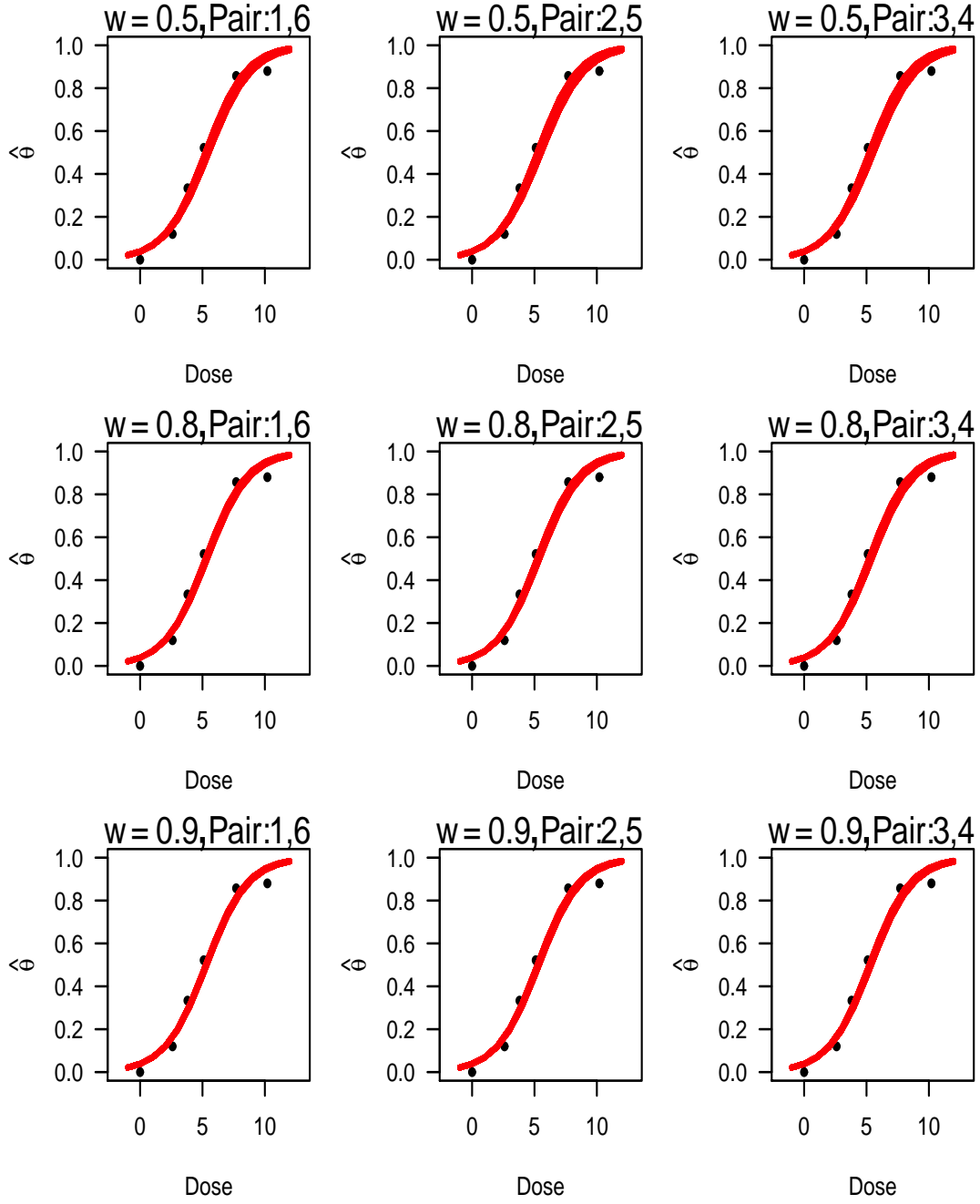


Figure D.1: Top: Plots of mixed ImpLogit model with $\nu = 1.5$, $w = 0.5$ and selected pairs θ_1, θ_6 (top-left), θ_2, θ_5 (top-middle), θ_3, θ_4 (top-right). Middle: $\nu = 1.5$, $w = 0.8$ and selected pairs θ_1, θ_6 (middle-left), θ_2, θ_5 (middle-middle) and θ_3, θ_4 (middle-right). Bottom: $\nu = 1.5$, $w = 0.9$ and selected pairs θ_1, θ_6 (bottom-left), θ_2, θ_5 (bottom-middle) and θ_3, θ_4 (bottom-right).

Table D.7: Interval estimates of β_0 and β_1 that correspond to top plots in Figure D.2 for 2-parameter ImpLogit model under IBLnM.

selected pair	$\hat{\beta}_0$	$\bar{\hat{\beta}}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\hat{\beta}}_1$	$\text{imp}(\hat{\beta}_1)$
$\theta_1, \theta_6(x_1, x_6)$	-3.225	-3.22	0.005	0.591	0.614	0.023
$\theta_2, \theta_5(x_2, x_5)$	-3.227	-3.223	0.004	0.585	0.614	0.029
$\theta_3, \theta_4(x_3, x_4)$	-3.227	-3.223	0.004	0.583	0.614	0.031

Table D.8: Interval estimates of β_0 and β_1 that correspond to middle plots in Figure D.2 for 2-parameter ImpLogit model under IBLnM.

selected pair	$\hat{\beta}_0$	$\bar{\hat{\beta}}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\hat{\beta}}_1$	$\text{imp}(\hat{\beta}_1)$
$\theta_1, \theta_6(x_1, x_6)$	-3.225	-3.219	0.006	0.594	0.622	0.028
$\theta_2, \theta_5(x_2, x_5)$	-3.228	-3.219	0.009	0.594	0.624	0.030
$\theta_3, \theta_4(x_3, x_4)$	-3.229	-3.218	0.011	0.590	0.623	0.033

Table D.9: Interval estimates of β_0 and β_1 that correspond to bottom plots in Figure D.2 for 2-parameter ImpLogit model under IBLnM.

selected pair	$\hat{\beta}_0$	$\bar{\hat{\beta}}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\hat{\beta}}_1$	$\text{imp}(\hat{\beta}_1)$
$\theta_1, \theta_6(x_1, x_6)$	-3.229	-3.220	0.009	0.595	0.623	0.028
$\theta_2, \theta_5(x_2, x_5)$	-3.229	-3.215	0.014	0.596	0.626	0.030
$\theta_3, \theta_4(x_3, x_4)$	-3.230	-3.216	0.014	0.591	0.626	0.035

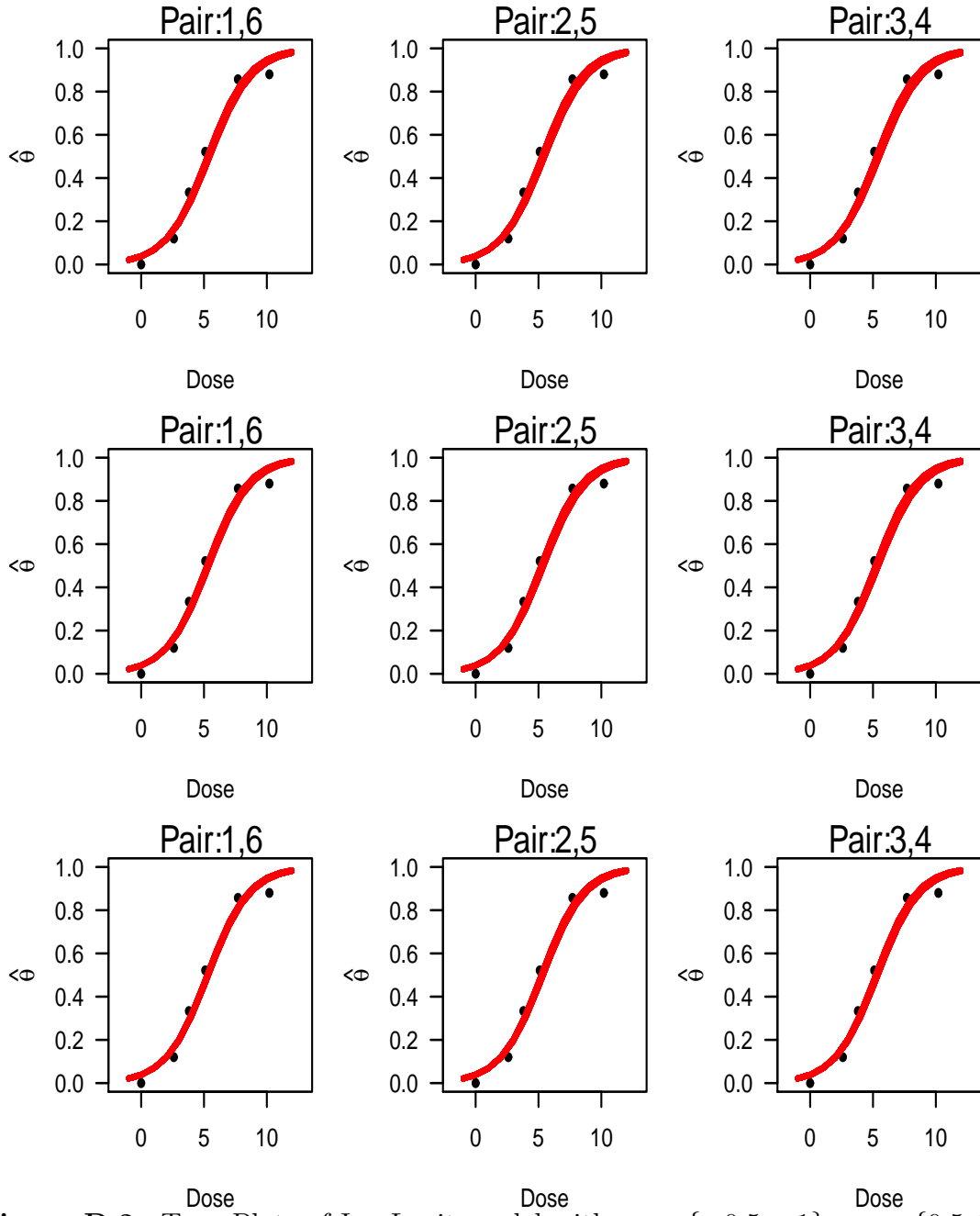


Figure D.2: Top: Plots of ImpLogit model with $\mu_j \in \{-0.5, -1\}$, $\mu_k \in \{0.5, 1\}$, $\sigma_j = \sigma_k \in \{0.55, 0.7, 0.85, 1, 1.15, 1.3, 1.45\}$ and selected pairs θ_1, θ_6 (top-left), θ_2, θ_5 (top-middle), θ_3, θ_4 (top-right). Middle: $\mu_j \in \{-0.5, -1\}$, $\mu_k \in \{0.5, 1\}$, $\sigma_j = \sigma_k \in \{0.3, 0.45, 0.6, 0.75, 0.9, 1.05, 1.2\}$. Bottom: $\mu_j \in \{-0.5, -1\}$, $\mu_k \in \{0.5, 1\}$, $\sigma_j = \sigma_k \in \{0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}$.

Table D.10: Interval estimates of β_0 and β_1 for 2-parameter increasing ImpLogit model under a small sample and when $\nu = 0.5$.

selected pair	$\hat{\beta}_0$	$\bar{\hat{\beta}}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\hat{\beta}}_1$	$\text{imp}(\hat{\beta}_1)$
$\theta_1, \theta_6(x_1, x_6)$	-3.246	-3.201	0.045	0.571	0.632	0.061
$\theta_2, \theta_5(x_2, x_5)$	-3.249	-3.207	0.042	0.568	0.633	0.065
$\theta_3, \theta_4(x_3, x_4)$	-3.248	-3.205	0.043	0.565	0.633	0.068

Table D.11: Interval estimates of β_0 and β_1 for 2-parameter increasing ImpLogit model under a small sample and when $\nu = 1$.

selected pair	$\hat{\beta}_0$	$\bar{\hat{\beta}}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\hat{\beta}}_1$	$\text{imp}(\hat{\beta}_1)$
$\theta_1, \theta_6(x_1, x_6)$	-3.246	-3.199	0.047	0.575	0.640	0.065
$\theta_2, \theta_5(x_2, x_5)$	-3.247	-3.198	0.049	0.571	0.642	0.071
$\theta_3, \theta_4(x_3, x_4)$	-3.248	-3.198	0.050	0.571	0.642	0.071

Table D.12: Interval estimates of β_0 and β_1 for 2-parameter increasing ImpLogit model under a small sample and when $\nu = 2$.

selected pair	$\hat{\beta}_0$	$\bar{\hat{\beta}}_0$	$\text{imp}(\hat{\beta}_0)$	$\hat{\beta}_1$	$\bar{\hat{\beta}}_1$	$\text{imp}(\hat{\beta}_1)$
$\theta_1, \theta_6(x_1, x_6)$	-3.247	-3.199	0.048	0.573	0.642	0.069
$\theta_2, \theta_5(x_2, x_5)$	-3.249	-3.197	0.052	0.572	0.645	0.073
$\theta_3, \theta_4(x_3, x_4)$	-3.250	-3.195	0.055	0.572	0.646	0.074

APPENDIX E

GRAPHS FROM CHAPTER 4

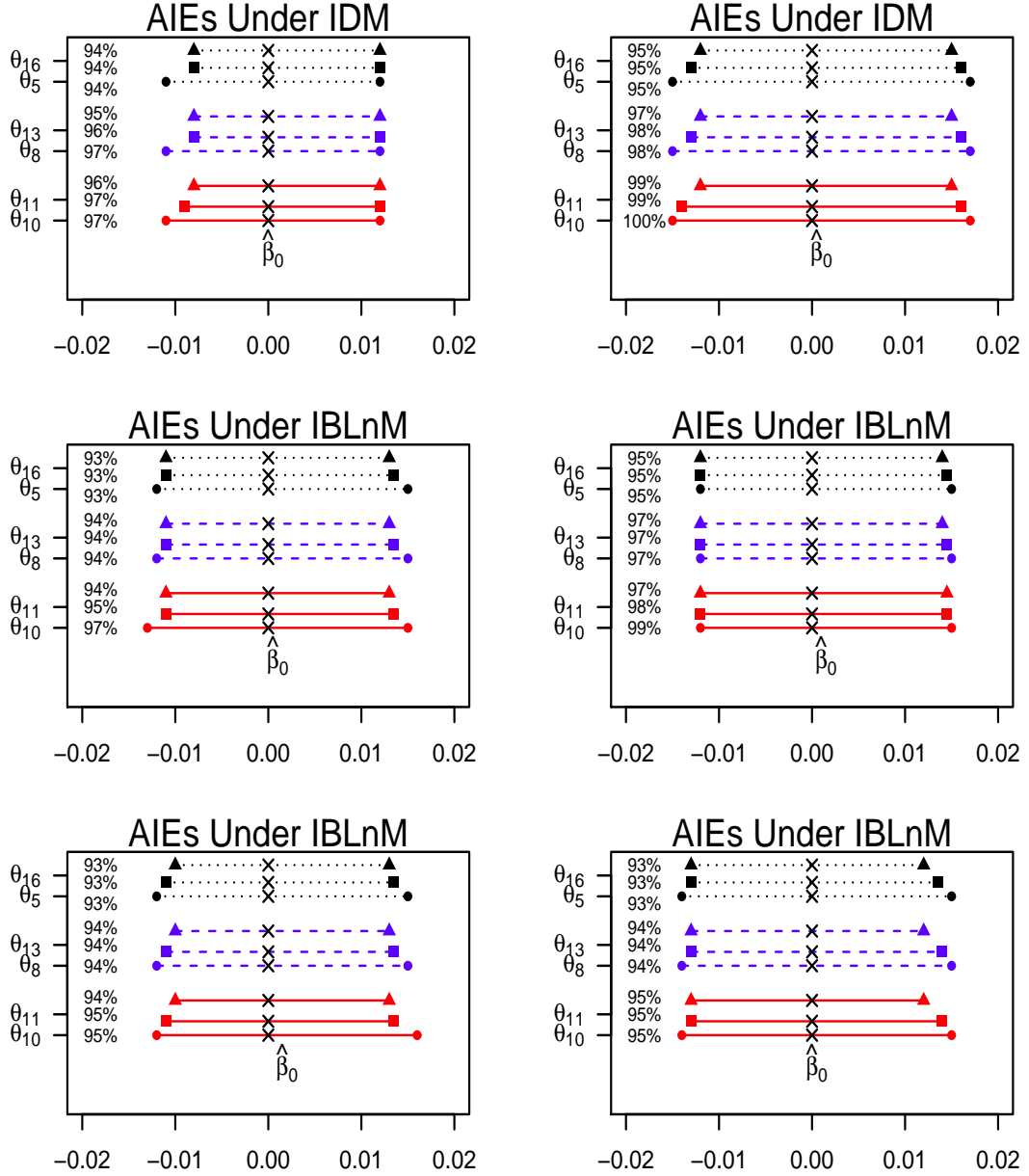


Figure E.1: Plots of AIEs of β_0 when $m = 20$ and $n_i = 2, \forall i$, and selections of θ_5 and θ_{16} , θ_8 and θ_{13} and θ_{10} and θ_{11} . Percentage values are for single interval estimates that include true β_0 .

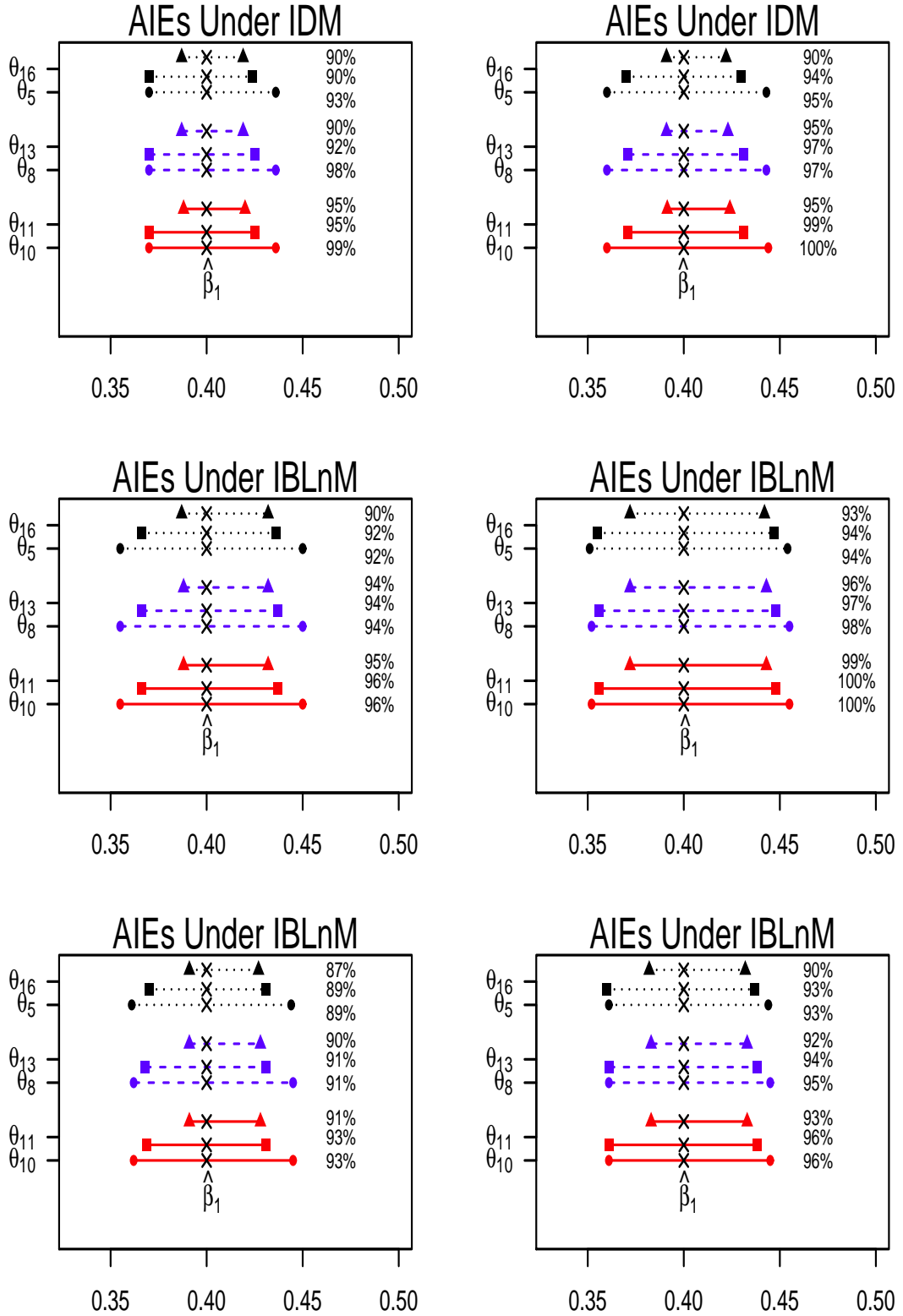


Figure E.2: Plots of AIEs of β_1 when $m = 20$ and $n_i = 2, \forall i$, and selections of θ_5 and θ_{16} , θ_8 and θ_{13} and θ_{10} and θ_{11} . Percentage values are for single interval estimates that include true β_1 .

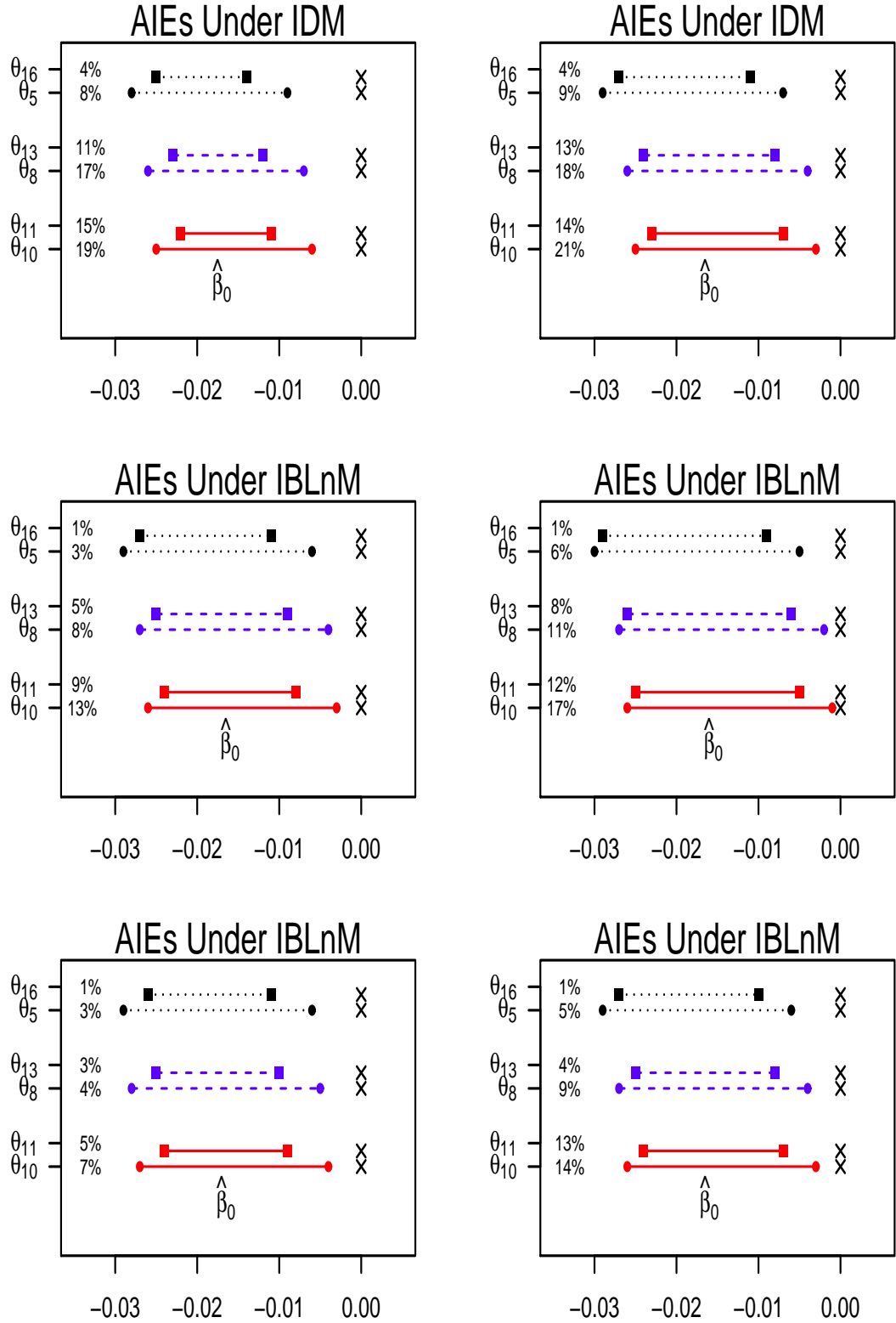


Figure E.3: Plots of AIEs of β_0 when $m = 20$ and $n_i = 2, \forall i$, and selections of θ_5 and θ_{16} , θ_8 and θ_{13} and θ_{10} and θ_{11} . Percentage values are for single interval estimates that include true β_0 .

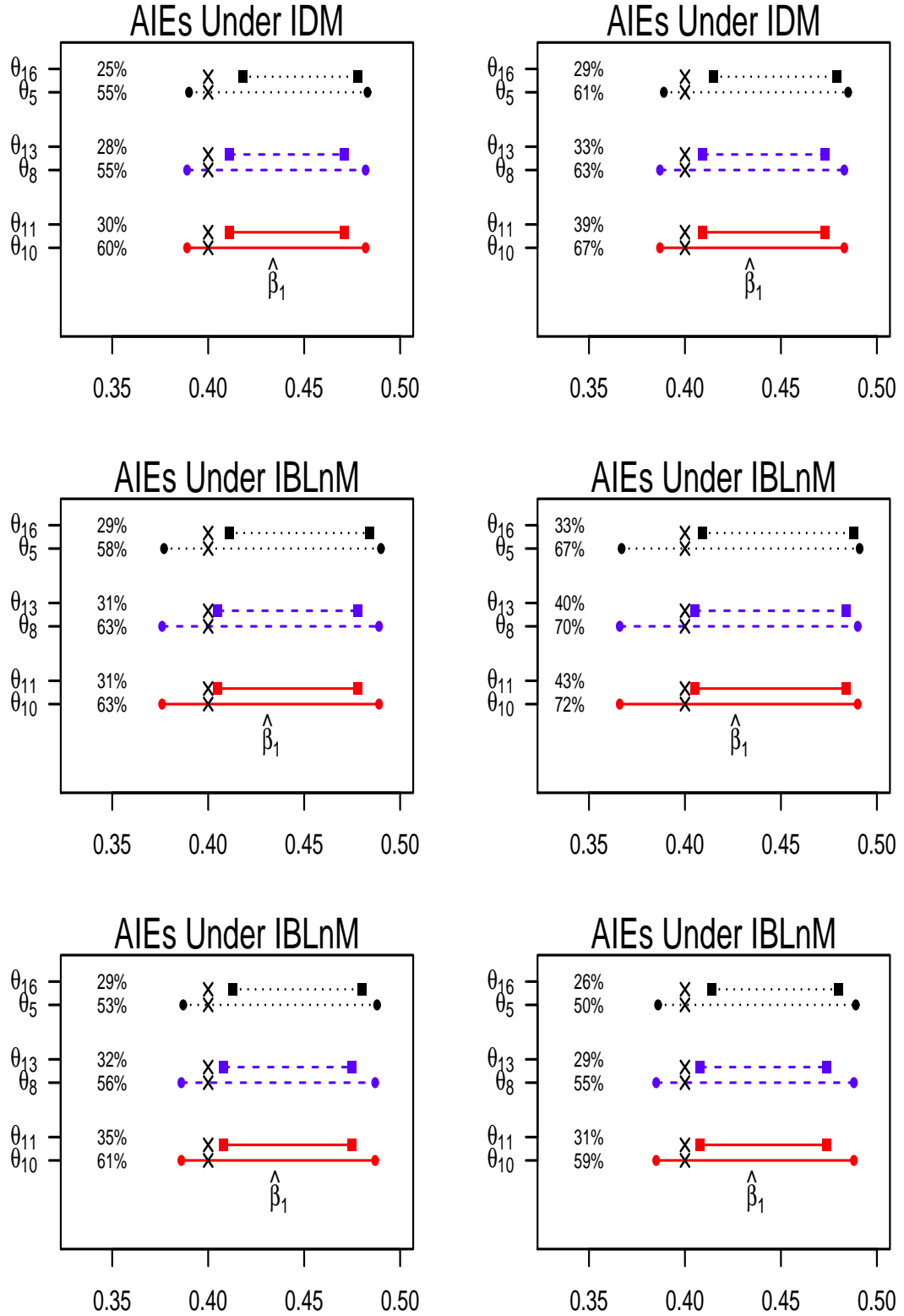


Figure E.4: Plots of AIEs of β_1 when $m = 20$ and $n_i = 2, \forall i$, and selections of θ_5 and θ_{16} , θ_8 and θ_{13} and θ_{10} and θ_{11} . Percentage values are for single interval estimates that include true β_1 .

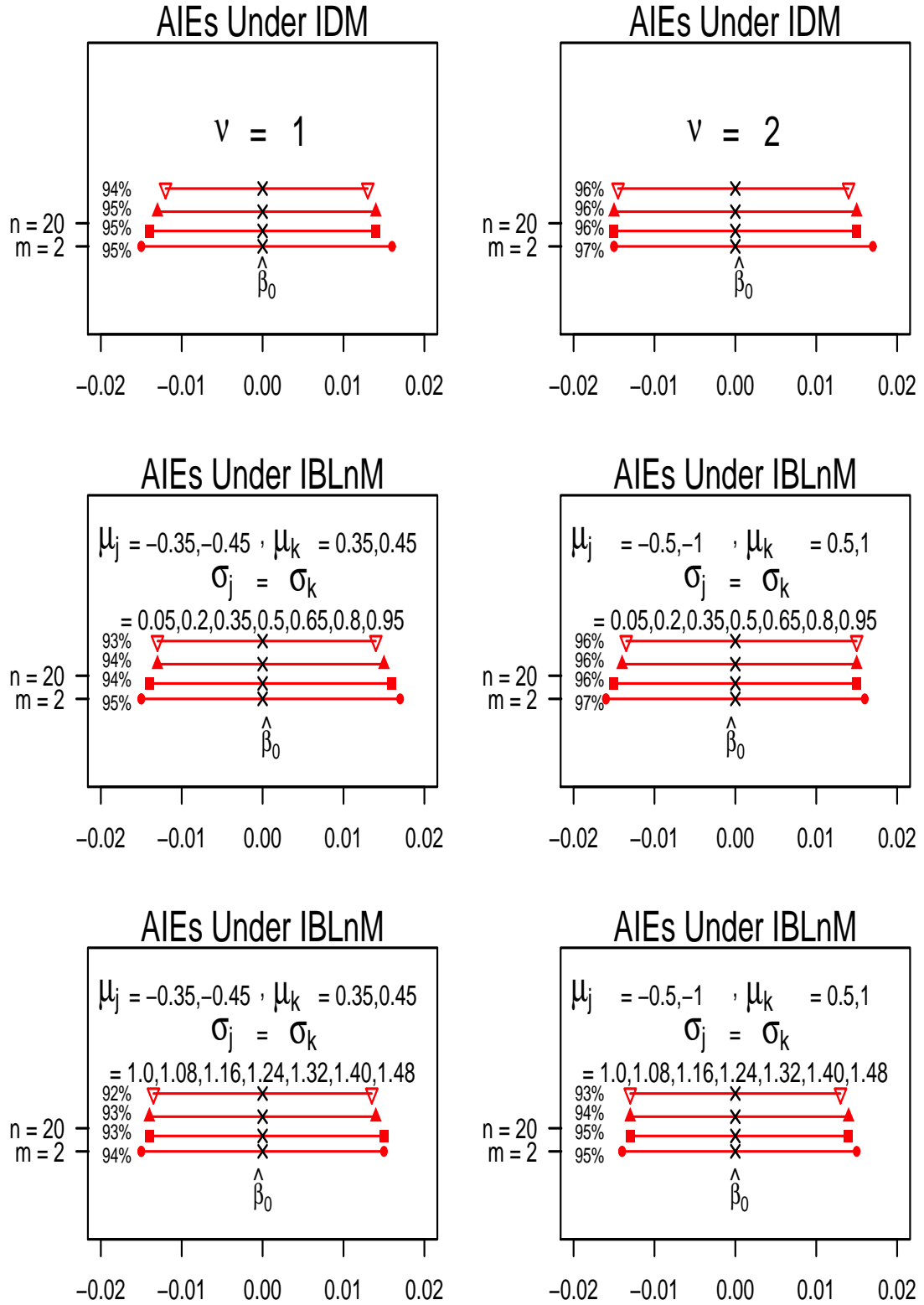


Figure E.5: Plots of AIEs of β_0 ended by \bullet when $x_1 = -0.5$ and $x_2 = 0.5$, \blacksquare when $x_1 = -1$ and $x_2 = 1$, \blacktriangle when $x_1 = -2.5$ and $x_2 = 2.5$ and ∇ when $x_1 = -3.85$ and $x_2 = 3.85$. Percentage values are for single interval estimates that include true β_0 .

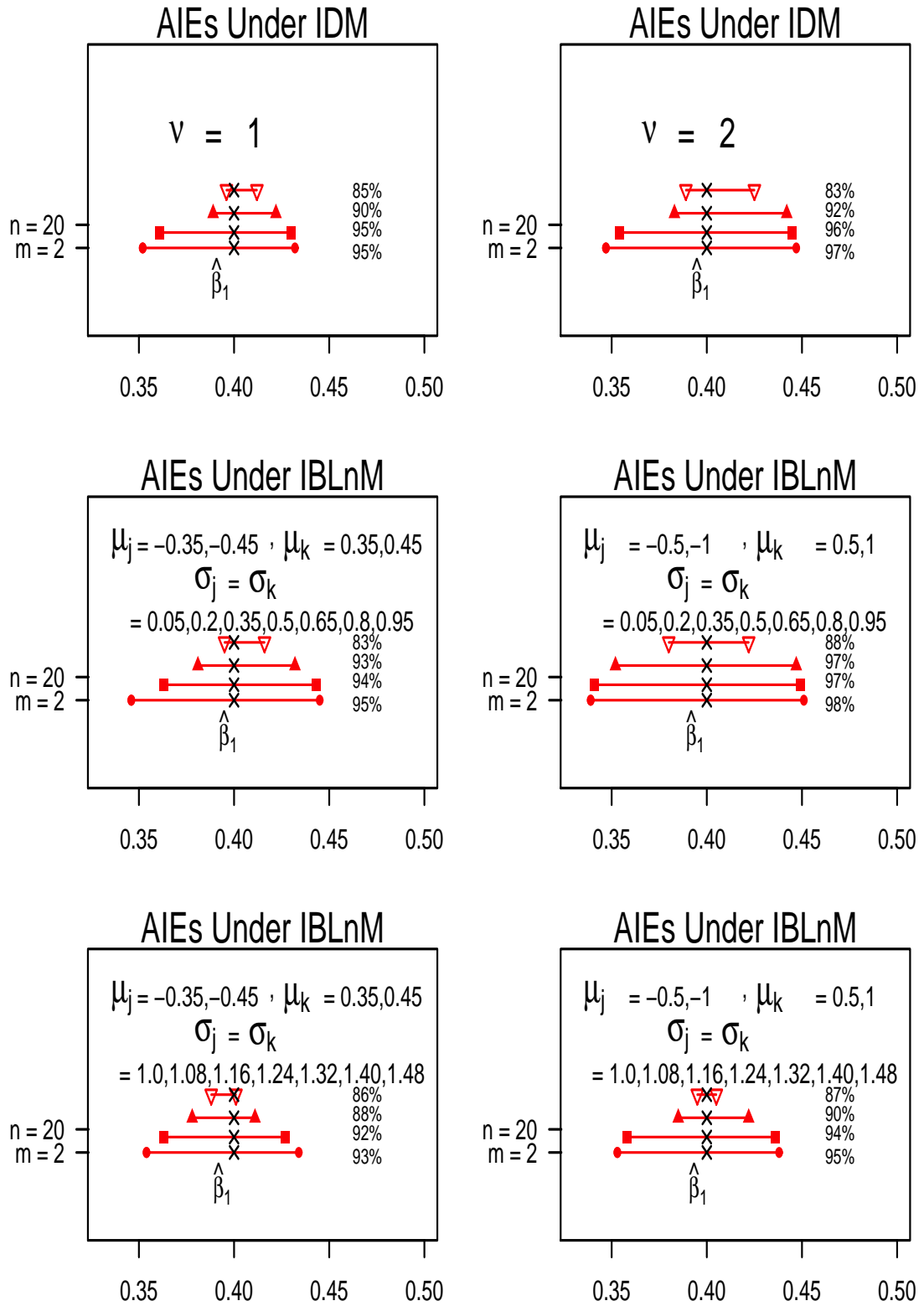


Figure E.6: Plots of AIEs of β_1 ended by ● when $x_1 = -0.5$ and $x_2 = 0.5$, ■ when $x_1 = -1$ and $x_2 = 1$, ▲ when $x_1 = -2.5$ and $x_2 = 2.5$ and ▽ when $x_1 = -3.85$ and $x_2 = 3.85$. Percentage values are for single interval estimates that include true β_1 .

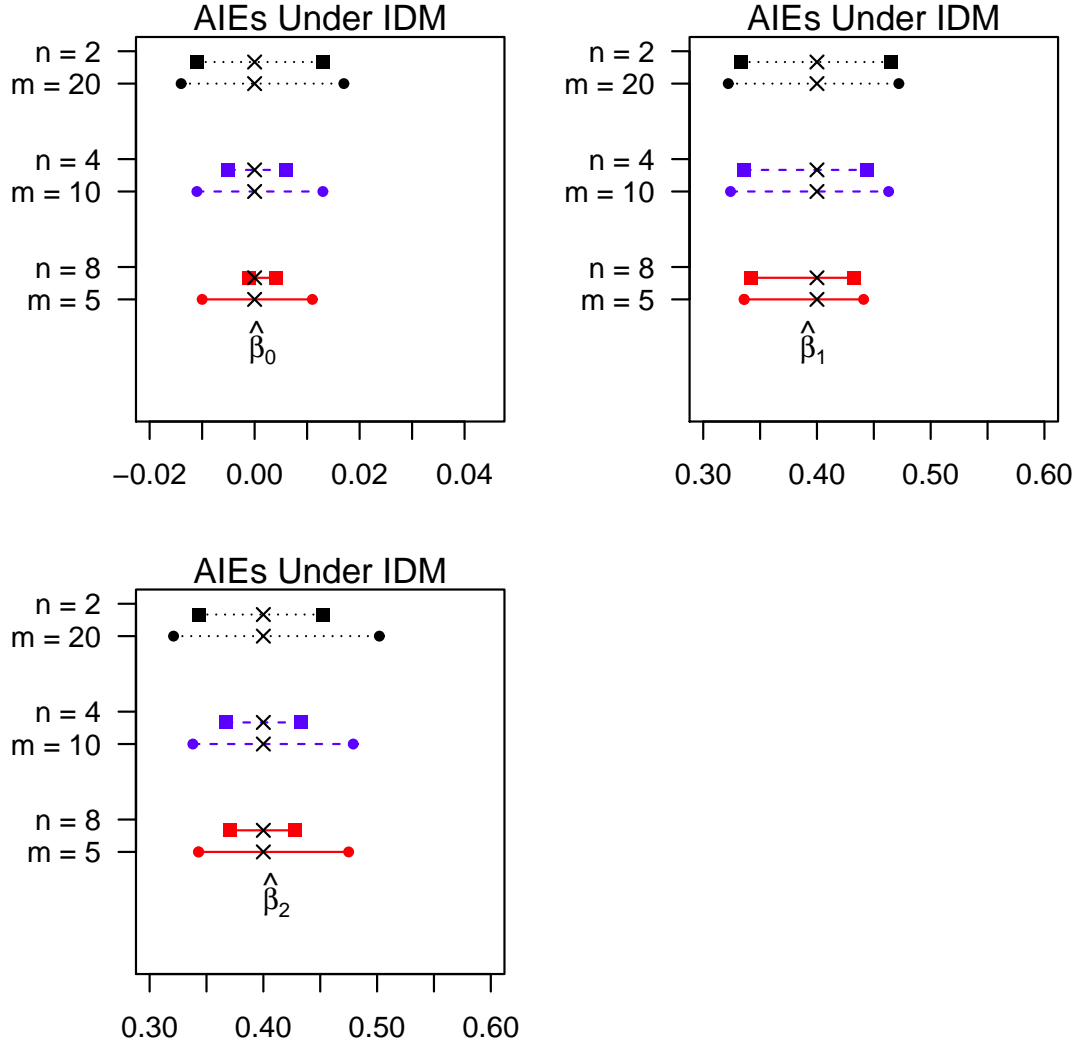


Figure E.7: Plots of AIEs in 3-parameter ImpLogit model under IDM with $\nu = 1$, of β_0 , β_1 and β_2 ended by \bullet when $x_{i1} \in [-1, 1]$ and $x_{i2} \in [-1, 1]$, \blacksquare when $x_{i1} \in [-2.5, 2.5]$ and $x_{i2} \in [-2.5, 2.5]$, $i = 1, \dots, m$.

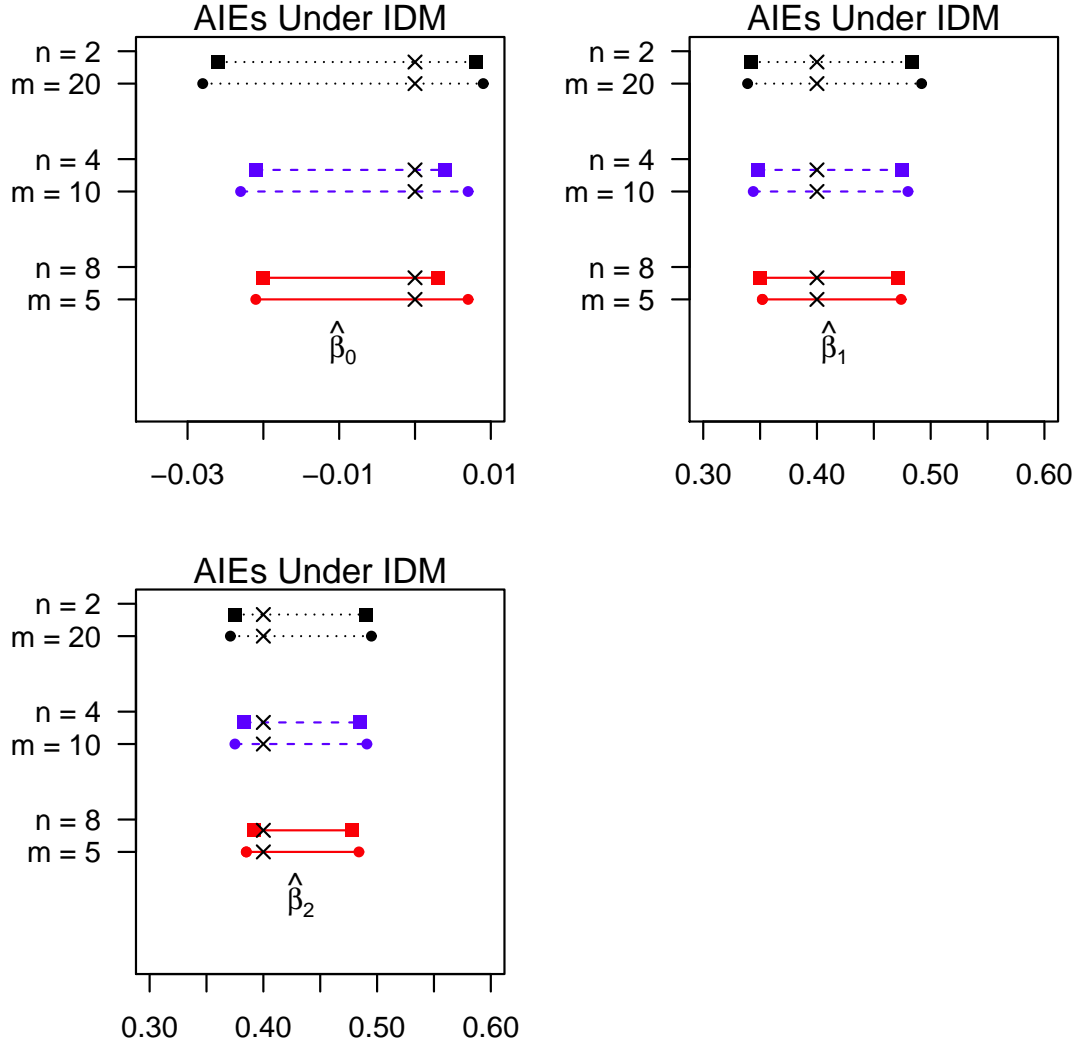


Figure E.8: Plots of AIEs in 3-parameter ImpLogit model under IDM with $\nu = 1$, of β_0 , β_1 and β_2 ended by • when $x_{i1} \in [-1, 1]$ and $x_{i2} \in [1, 3]$, ■ when $x_{i1} \in [-2.5, 2.5]$ and $x_{i2} \in [1, 6]$, $i = 1, \dots, m$.

APPENDIX F

FISHER INFORMATION MATRIX FOR LOGISTIC REGRESSION

To find the information matrix, the likelihood function of the logistic regression model is

$$\begin{aligned}
 L(\theta_1, \dots, \theta_m | \mathbf{y}) &= \prod_{i=1}^m \left[\binom{n_i}{y_i} \theta_i^{y_i} (1 - \theta_i)^{n_i - y_i} \right] \\
 \implies \log [L(\theta_1, \dots, \theta_m | \mathbf{y})] &= \sum_{i=1}^m \log \binom{n_i}{y_i} + \sum_{i=1}^m y_i \log \theta_i + \sum_{i=1}^m (n_i - y_i) \log (1 - \theta_i) \\
 \implies \log [L(\beta_0, \beta_1 | \mathbf{y})] &= \sum_{i=1}^m \log \binom{n_i}{y_i} + \sum_{i=1}^m y_i \log \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) \\
 &\quad + \sum_{i=1}^m (n_i - y_i) \log \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right). \tag{F.1}
 \end{aligned}$$

Then (F.1) is used in the information matrix of β_0 and β_1 as

$$I(\beta_0, \beta_1) = \begin{bmatrix} -E \left[\frac{\partial^2}{\partial \beta_0^2} \log L(\beta_0, \beta_1 | \mathbf{y}) \right] & -E \left[\frac{\partial^2}{\partial \beta_0 \partial \beta_1} \log L(\beta_0, \beta_1 | \mathbf{y}) \right] \\ -E \left[\frac{\partial^2}{\partial \beta_0 \partial \beta_1} \log L(\beta_0, \beta_1 | \mathbf{y}) \right] & -E \left[\frac{\partial^2}{\partial \beta_1^2} \log L(\beta_0, \beta_1 | \mathbf{y}) \right] \end{bmatrix}, \tag{F.2}$$

where from (F.2),

$$\begin{aligned}
 &\frac{\partial^2}{\partial \beta_0^2} \log L(\beta_0, \beta_1 | \mathbf{y}) \\
 &= \frac{\partial^2}{\partial \beta_0^2} \left[\sum_{i=1}^m y_i \log \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) \right] + \frac{\partial^2}{\partial \beta_0^2} \left[\sum_{i=1}^m (n_i - y_i) \log \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right) \right] \\
 &= \frac{\partial}{\partial \beta_0} \left\{ \sum_{i=1}^m y_i \frac{\frac{\partial}{\partial \beta_0} \left[\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right]}{\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}} \right\} + \frac{\partial}{\partial \beta_0} \left\{ \sum_{i=1}^m (n_i - y_i) \frac{\frac{\partial}{\partial \beta_0} \left[\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right]}{\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}}} \right\}
 \end{aligned}$$

$$\begin{aligned}
&= \frac{\partial}{\partial \beta_0} \left\{ \sum_{i=1}^m y_i \frac{\frac{(1 + e^{\beta_0 + \beta_1 x_i}) e^{\beta_0 + \beta_1 x_i} - (e^{\beta_0 + \beta_1 x_i})^2}{(1 + e^{\beta_0 + \beta_1 x_i})^2}}{\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}} \right\} \\
&+ \frac{\partial}{\partial \beta_0} \left\{ \sum_{i=1}^m (n_i - y_i) \frac{\frac{-e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2}}{\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}}} \right\} \\
&= \frac{\partial}{\partial \beta_0} \left\{ \sum_{i=1}^m y_i \frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right\} + \frac{\partial}{\partial \beta_0} \left\{ \sum_{i=1}^m (n_i - y_i) \frac{-e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right\} \\
&= \sum_{i=1}^m y_i \frac{-e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} + \sum_{i=1}^m (n_i - y_i) \frac{(1 + e^{\beta_0 + \beta_1 x_i}) (-e^{\beta_0 + \beta_1 x_i}) + (e^{\beta_0 + \beta_1 x_i})^2}{(1 + e^{\beta_0 + \beta_1 x_i})^2} \\
&= \sum_{i=1}^m y_i \frac{-e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} + \sum_{i=1}^m (n_i - y_i) \frac{-e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} = \sum_{i=1}^m n_i \frac{-e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} \\
&= - \sum_{i=1}^m n_i \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right) = - \sum_{i=1}^m n_i \theta_i (1 - \theta_i). \tag{F.3}
\end{aligned}$$

The second derivative with respect to β_1 is

$$\frac{\partial^2}{\partial \beta_1^2} \log L(\beta_0, \beta_1 | \mathbf{y}) = - \sum_{i=1}^m n_i x_i^2 \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right) = - \sum_{i=1}^m n_i x_i^2 \theta_i (1 - \theta_i), \tag{F.4}$$

and

$$\begin{aligned}
\frac{\partial^2}{\partial \beta_0 \partial \beta_1} \log L(\beta_0, \beta_1 | \mathbf{y}) &= \frac{\partial^2}{\partial \beta_1 \partial \beta_0} \log L(\beta_0, \beta_1 | \mathbf{y}) \\
&= - \sum_{i=1}^m x_i n_i \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right) = - \sum_{i=1}^m x_i n_i \theta_i (1 - \theta_i), \tag{F.5}
\end{aligned}$$

then plugging (F.3), (F.4) and (F.5) in (F.2) gives FIM as

$$\begin{bmatrix} \sum_{i=1}^m n_i \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right) & \sum_{i=1}^m x_i n_i \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right) \\ \sum_{i=1}^m x_i n_i \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right) & \sum_{i=1}^m n_i x_i^2 \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \left(\frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right) \end{bmatrix}. \tag{F.6}$$